https://ntrs.nasa.gov/search.jsp?R=19780015029 2020-03-22T05:10:29+00:00Z

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

SD78-MSFC-2204

SD78-MSFC-2204

Technical Report

DATA SYSTEMS ELEMENTS TECHNOLOGY ASSESSMENT AND SYSTEM SPECIFICATIONS (ISSUE NO. 2)

(NASA-CR-150677) DATA SYSTEMS ELEMENTS TECHNOLOGY ASSESSMENT AND SYSTEM SPECIFICATIONS, ISSUE NO. 2 (Teledyne Brown Engineering) 758 p HC A99/MF A01 CSCL 05B

N78-22972

G3/85 Unclas 16644

April 1978

TELEDYNE BROWN ENGINEERING

Cummings Research Park • Huntsville, Alabama 35807

TECHNICAL REPORT
SD78-MSFC-2204

DATA SYSTEMS ELEMENTS TECHNOLOGY ASSESSMENT AND SYSTEM SPECIFICATIONS (ISSUE NO. 2)

April 1978

Prepared For

DATA SYSTEMS LABORATORY
GEORGE C. MARSHALL SPACE FLIGHT CENTER
MARSHALL SPACE FLIGHT CENTER, ALABAMA

Contract No. NAS8-32539

Prepared By

SYSTEMS DIVISION
TELEDYNE BROWN ENGINEERING
HUNTSVILLE, ALABAMA

ABSTRACT

The ability to satisfy the objectives of future NASA Office of Applications programs is dependent on technology advances in a number of areas of data systems. This report examines the hardware and software technology of end-to-end data systems (data processing elements through ground processing, dissemination, and presentation) in terms of state of the art, trends, and projected developments in the 1980 to 1985 timeframe. Capability is considered in terms of elements that are either commercially available or that can be implemented from commercially available components with minimal development.

G. G. Hodges
Program Manager

APPROVED:

R. E. Alger
Manager
Data Processing Department

PRECEDING PAGE BLANK NOT FILMED

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Concluded

ACKNOWLEDGEMENTS

Many authorities in the field of data systems responded to requests for information in support of this report via the mails, telecom, and visits. Significant parts of the forecasts are the result of their contributions. Many of the respondents generously gave permission to use their names as sources of the information provided. However, the sponsoring agency (NASA) requested that specific references be made only to published literature to eliminate the possibility of embarassment to either individuals or organizations. Individuals and organizations that contributed are listed alphabetically at the end of the report. Their contributions are gratefully acknowledged.

A number of individuals within Teledyne Brown Engineering contributed to the various report sections and subsections. These authors are listed at the end of the report. The technical editors within the TBE publications department also deserve credit for their assistance in editing and coordinating the production of this document.

In addition to the contributors and authors, two individuals deserve special credit for their efforts in the guidance and preparation of this report. The first is Mr. Olen Ely, the program manager during the initial phases of the contract. Also, Mr. Ken Williamson, the NASA technical coordinator, contributed significantly in terms of guidance and helpful suggestions.

EXECUTIVE SUMMARY

INTRODUCTION

The ability to satisfy the objectives of future NASA Office of Applications (OA) programs depends on technology advances in all areas of data systems. This report examines the technology of end-to-end data systems (space processing elements through ground processing, dissemination, and presentation) in terms of state of the art, trends, and projected developments in the 1980 to 1985 timeframe. Capability is considered in terms of elements that are either commercially available or can be implemented from commercially available components with minimal development. The following paragraphs of this Executive Summary highlight the trends and projected developments in some of the more important areas of data systems as far as NASA's future programs are concerned. The summary and report itself begin with a technology perspective that focuses on the ongoing revolution in microelectronics. This first section of the report presents a background and a basis for many of the trends and projections that are given in the remaining sections of the report.

Data systems technology hardware is advancing at an unprecedented rate. Announcements of new products that outperform existing systems are made daily. The majority of advances are primarily the result of improvements in integrated circuit technology. Products are becoming faster, smaller, less expensive, more efficient, more flexible, and more reliable, to name a few areas of improvement. Although the major advances are occurring at the chip level, these advances are being implemented at the system level almost as rapidly as they occur. This report presents the technology status at both the chip level and the system level, with emphasis on the latter area since this is where the effects of NASA's programs will be most heavily felt.

Although the major emphasis throughout the report is on hardware technology, a section on software technology has been added to this report update. Software technology is addressed in Section 11. Software is

important not only because it is a potential bottleneck in data system development but also because it is expected to represent an increasingly larger percentage of the NASA budget for future data systems. Of course this is the case only because of the significant advances that are occurring in hardware technology.

The Executive Summary is presented in the same sequence as the report. The emphasis throughout this summary is on trends and projected developments.

TECHNOLOGY PERSPECTIVE

Microelectronic technology has advanced to the point where many established data system concepts are already being reevaluated in order to take advantage of these advances. Although past developments in microelectronic technology have already significantly reduced the cost per unit of performance of data system hardware, the impact of future developments should be even greater.

The state of the art in large-scale integrated (LSI) logic device technology includes logic devices with compexities in excess of 2×10^4 transistors per chip and memory devices with complexities in excess of 1.3×10^5 transistors per chip. As complex as these devices may seem, submicron technology promises to add approximately two orders of magnitude to the complexity of logic and memory devices by 1985. Currently, the two major candidates for the manufacture of suboptical microelectronic devices are X-ray lithography and electron-beam fabrication. However, photolithography continues to improve and push into the suboptical domain with the use of deep ultraviolet lithography. For example, one company recently reported a photolithographic system capable of fabricating 0.5-μm linewidths.

In addition to advancements in wafer processing, considerable work is being done toward developing new logic families that will provide improvements in both gate delay times and speed-power products. Additionally, new semiconducting materials, such as gallium ansenide, that

have higher carrier mobility (hence higher speed) than silicon are already in the advanced stages of development. During the next decade, impressive data system performance improvements will occur as the advances in submicron processing techniques are combined with new device technologies to produce very large integrated circuitry that offers significant increases in complexity and performance over today's technology.

SPACE PROCESSING ELEMENTS

Space processing elements have traditionally been characterized by extremely rugged construction, relatively small size, low-power operation, high reliability, and high cost. In order to achieve the required reliability and performance, a considerable effort was normally expended on packaging design, efficient electrical design, analytical models (worst-case electrical, thermal, etc) and extensive quality controls. The design of many early spaceborne computers (with a few exceptions) did not seriously consider either software development or architectural compatibility with already existing machines.

Developers of state-of-the-art spaceborne computers such as the NSSC II, which is currently being developed by IBM for MSFC, recognizes the importance of architectural compatibility and the rising cost of software; therefore, the NSSC II utilizes an IBM 360 architecture.

By 1985, commercial minicomputers with approximately the same complexity as the NSSC II will be available on one or two chips. The availability of such a great amount of computing power in such a small package will lead to greater application of imbedded processors and distributed processing within spacecraft subsystems. Because of the extremely small size of commercial minicomputers, there will be less difference between the physical configuration of a spaceborne computer and a commercial minicomputer and essentially the same machines may be used in both applications.

SPACE DATA STORAGE ELEMENTS

Space data storage has typically been performed with magnetic tape recorders that are limited in capacity, reliability, and lifetime. Although magnetic tape recorders will still be used extensively in space into the mid-1980's and possibly beyond, new technologies that are potentially more reliable are beginning to emerge. The most promising technology for replacing tape recorders across a broad front in space is magnetic bubbles.

NASA is currently developing a prototype 10^8-bit bubble memory system. Although this system will not be competitive with existing magnetic tape recorders in terms of volume, weight, power, and storage capacity, it will provide for an early demonstration of flyable bubble memory recorder. Additionally, it will provide a mechanism for demonstrating the alleged higher reliability of bubble memory storage systems.

The potential for bubble memory spaceborne data recorders lies in the expected chip storage capacity increase from the currently available 10^5-bit chips to 10^8-bit chips by 1985. The larger-capacity bubble chips will make the performance of bubble memory systems competitive with tape data recorders for storage capacities of up to 10^{10} bits by 1985.

High density airborne digital tape recorders can be generally classified as rotary-head recorders and fixed-head recorders. The fixed-head recorders currently offer the highest data transfer rates. RCA is developing a 420-Mbit/sec fixed-head data recorder for GSFC that records 120 active data channels on a 2-in.-wide tape.

Authorities are predicting that the rotary-head tape recorders will be used extensively in future data systems that either require very high storage capacities or must interface with onboard digital computers that store and access data in blocks, rather than continuously as is currently done in most aerospace recorder systems.

Technology projections for 1985 are for aerospace data recorders with storage densities of up to 10^7 bits/in^2 and data transfer rates of up to 10^9 bits/sec.

SPACE DATA HANDLING ELEMENTS

Future general-purpose data systems will provide increased flexibility, reliability, and data throughput capacity. Microcomputers will be incorporated into future remote terminal unit designs to provide remote programming and processing capabilities that will significantly increase the overall data system throughput capacity by eliminating much of the nonessential data bus traffic and reducing the central computer processing load. Space-qualified 64-Kbit RAM chips and 1-Mbit bubble memory chips should be available by 1980. These devices will make it feasible to package microcomputers with upwards of 10^6 bytes of memory by 1980 without increasing the current size of the remote unit. Improvements in analog and digital circuit integration levels will more than offset the additional size and power required for the remote unit processor and memory functions.

By 1980, low-cost, low-power, monolithic analog-to-digital converters and digital-to-analog converters should be available from a number of different sources.

Technology projections for 1985 that will impact the ultra-high data rate and high-speed digital data handling systems of the future include the following:

- Space-qualified GaAs MESFET logic with propagation delays of less than 100 psec and clock rates in excess of 5 GHz will be available.
- All monolithic analog-to-digital series-parallel converters will be available with a conversion speed of 10 nsec for 8 bits.
- Hybrid analog-to-digital parallel converters will be available with a conversion speed of 2 nsec for 8 bits.
- Short-channel MOS analog switches will be available with switching speeds less than 500 psec.

These developments will provide the technology neccessary to build data systems having analog sample rates of up to 400 MHz for 8-bit resolution and output PCM data rates of up to 5 GHz for a single output serial data stream. This technology will support the development of real-time imaging systems similar to the Thematic Mapper with resolutions of less than 5 m.

Projected power consumption per gate for the high-power 5-GHz GaAs MESFET logic is 10 mW per gate. At this power level, only medium-scale integration can be used. For applications that require logic speeds less than 1 GHz, low-power GaAs MESFET logic with power dissipations of only 0.3 mW per gate can be used to provide highly reliable LSI designs with integration levels in excess of 1,000 gates per chip.

These developments in high-speed, relatively low-power logic, coupled with the developments in bubble memories discussed in Section 3.2 and semiconductor memories discussed in Subsection 7.2.1, will make high-rate data handling much easier by 1985.

Another development that is likely to have a significant impact on data handling system designs by 1980 is the optical data bus. Both the Air Force and the Navy are actively pursuing the development of optical transmission lines in an operational airborne environment. Advantages claimed include transmitting higher serial data rates (up to 100 Mbits/sec), reduced weight, and less susceptibility to RFI/EMI/EMP.

SPACE-TO-GROUND COMMUNICATIONS ELEMENTS

Both solid-state and tube-type power sources for space-to-ground communication systems are currently experiencing improvements in power output, efficiency, and size over a wide range of frequency bands, and the current trends are expected to continue through the 1980 to 1985 time period. Transmitters will be using a variety of devices for different frequency bands between now and 1985. Below 6 GHz, bipolar devices will be competing with TWTs for size, power, and efficiency. FETs should be equivalent to the IMPATTs of today in the 6- to 50-GHz band by the mid-1980's, whereas IMPATTs should improve in power and efficiency to provide approximately 1 to 2 W at 100 GHz. Efficiency in FETs should increase significantly by 1985, approaching 70% of the theoretical maximum. In spite of these advances, TWTs will continue to either equal or outperform solid-state devices in terms of efficiency, gain, and power output in most frequency bands.

The major areas of research in communications are going on at frequencies above 30 GHz, including the optical band. Improvements in these frequency bands are likely to be reflected in the lower frequency bands also.

Reflector-type antennas will be prevalent in all bands through 1985. These will be folding types below 10 GHz, and will be deployable at all frequencies. Prime focus and offset reflectors with cluster array feeds will be common. Phased-array antenna systems will be capable of competing with reflector types in performance, weight, and volume. However, they will be more expensive by a factor of 1.5 to 2. Phase arrays will handle much higher power without experiencing multipactor. Array feeds of low-loss air-stripline or microstrip networks using high-permittivity substrates will allow designs with sidelobes that are down more than 30 dB.

The use of the TDRSS for a civilian space data relay satellite will continue through 1990. An upgrading of the system would be technically feasible during the mid-1980's, but current NASA planning does not incorporate any design changes. The use of relay satellites employing laser technology will be feasible by 1985, with such systems capable of data rates in excess of 1 Gbit/sec.

Future developments in ground station antennas will be related to the continued increase in frequency of the satellite communication systems. Ground station antennas at these higher frequencies will use the same design techniques that have been developed for the lower bands. The ground antennas at the 4- to 6-GHz and 11- to 14-GHz bands will employ frequency reuse techniques that will require stricter pointing tolerances and lower sidelobe levels.

As a result of developments in super components, receiver sizes in all frequency bands will decrease to somewhere in the range of 25 to 50% of their present size by 1985, while performance will improve. Power consumption will continue to decrease to approximately 50% of the present consumption by 1985, especially for the higher-frequency receivers.

Solid-state devices will be capable of operating at higher frequencies with sufficiently low noise figures and high enough gains to replace TWTs in most systems.

The analog portion of the receiver of the future will shrink as more and more digital circuitry is used. The trend is to have digital circuitry from the conversion to baseband to the output of the receiver. When short (<512-bit) PN Codes are used, it is likely that CCD transversal filters will be employed.

Modulation schemes for data transmission through 1985 will be for either low-data-rate narrowband systems or high-data-rate wideband systems. The high-rate systems will use M-ary PN schemes for frequencies below the optical band. Data rates for both the optical and RF bands will range up to 1-Gbit/sec for terrestrial systems and up to 300-Mbits/sec for spaceborne systems.

Error correction coding will exhibit a trend toward convolutional codes for error correction and concatenated codes for multipurpose applications and extremely high-quality data. Future systems that use higher frequencies will be more susceptible to fading and atmospheric disturbances and will not necessarily employ the same error correction techniques used at the lower frequencies.

PREPROCESSING ELEMENTS

Preprocessing elements include those hardware and software elements that take the outputs from the ground communications system (ground station receivers) and perform the functions necessary to input the data into the ground processing system. Functions performed by the preprocessing elements include data detection and synchronization, demultiplexing, reformatting, data buffering, decoding, error correction, and calibration.

The TDRSS ground station, currently under development, represents the state of the art in operational ground stations. When operational in 1980, the TDRSS ground station will handle data rates to 300 Mbits/sec

and will employ a Viterbi decoder system that is capable of data rates to 50 Mbits/sec. Development hardware for future system applications is currently being tested in the laboratory with bit synchronizers that operate at 500 Mbits/sec and demultiplexers capable of handling rates to 1 Gbits/sec.

In general, the trends in preprocessing elements are toward higher data rates and more real-time processing and data handling, including real-time data reformatting, retransmission, and distribution.

Tape recorders currently used for ground data storage are mostly the fixed-head, high-density digital instrumentation type. Trends in high-density digital instrumentation recorders are toward wider tapes, more heads per inch of tape, and higher data rates. Several companies are currently developing fixed-head instrumentation recorders with data recording rates in excess of 200 Mbits/sec.

The use of rotary-head type data recorders will increase because of the higher data packing densities and faster start-stop times that can be achieved with the rotary-head systems.

Flexible high-rate instrumentation recorders should be available by 1985 that record and play back at data rates up to 1 Gbits/sec with data storage densities of 10^7 bits/in^2.

GROUND PROCESSING ELEMENTS

Data processing elements covered in this report include microcomputers, minicomputers, large-scale computers, super-scale computers, bipolar memory systems, MOS memory systems, CCD memory systems, electron-beam memory systems, core memory, disks, tapes, and firmware. Software is discussed in Sections 8 and 11.

The market size and capabilities of microcomputers will continue to increase during the 1980 to 1985 timeframe. By providing very inexpensive low-end processing, microcomputers will find their way into many new designs, performing functions that could not be done economically with

programming (or done at all) before. The most powerful microcomputers of 1980 and 1985, capable of achieving instruction rates of 5 and 20 million instructions per second (MIPS), respectively, will expand the market for these computers on the high end. Software for this class of computers will become somewhat more sophisticated, but, with rare exception, will not rival that of larger machines.

The rapid growth in the capabilities of minicomputers experienced during the past decade will continue during the 1980 to 1985 timeframe. By taking advantage of the predicted improvements in LSI logic circuits and semiconductor memories, the most powerful minicomputers of 1980 and 1985 should be capable of 12 and 60 MIPS, respectively, at costs in the vicinity of $40,000. Mean time between failure for these systems is expected to approach 10,000 hr, and software offered with the machines is going to become much more sophisticated, rivaling that of medium-scale computers of today.

The historic 12 to 15% annual price/performance improvement in large-scale computers will continue during the 1980 to 1985 timeframe as a result of achievements in logic circuits, main memories, and mass storage. The desire for ever greater speed will force the continued growth of multilevel hierarchical storage systems. In particular, the use of buffer memories will expand. More emphasis will be placed on system reliability, with improvements achieved through the use of error-correcting codes in memory systems, fault-tolerant design, and hardware redundancy. The most powerful large-scale computers of 1980 will be capable of 45 MIPS, with increases to 100 MIPS by 1985. Expanded use of distributed data processing systems will occur in conjunction with, rather than instead of, the growth of large-scale computers. Many of the large-scale computers of the next decade will be geared to large data base management operations.

Super-scale computers of the next decade will continue to have a very limited market, addressing problems that throughput large amounts of data and in which computations are highly parallel in nature. Their

characteristics will continue to improve, with emphasis placed on larger memories and faster execution times since the quality of the solution to the problems addressed depends on the number of data points considered and on the number of computations performed. Both vector and parallel designs will be used to build systems.

Vector super-scale computers are expected to achieve 100 to 150 MIPS by the early 1980's, with increases to 200 to 300 MIPS by the mid-1980's. Parallel super-scale computers will be capable of achieving instruction rates as high as 1,000 MIPS by 1985. One limiting factor in achieving these rates is the extent to which the parallel or vector capabilities are utilized, which depends primarily on software advances. Since software development is slower than hardware development, it will be the limiting factors, preventing the more optimistic projections made by some experts from being achieved. Only very limited software support is available on most state-of-the-art super-scale computers. Major advances in software suppport are needed so that systems can be more fully utilized. For this reason, emphasis will be placed on software development during the next decade.

MOS technology, particularly NMOS, will continue to be the dominant processor memory technology during the 1980 to 1985 timeframe. Major improvements will occur in all performance categories of NMOS and CMOS memory systems during this time period. By 1980, chip densities will quadruple, and access and cycle times will improve by 40%. By 1985, chip densities will increase by an additional factor of 16, and access and cycle times will improve by 60%.

Charge Coupled Device (CCD) memories offer the greatest densities available in semiconductor read-write memories; however, CCD memories are generally organized as serial rather than random-access devices and thus have relatively long access times. This position should be maintained or improved on in the 1980 to 1985 timeframe, with devices storing 256 Kbits and 4 Mbits being available in 1980 and 1985, respectively.

Electron beam memories offer great potential in terms of performance, cost per bit, and total system capacity. However, for major advances to occur, the following problem areas must be overcome:

- Electron optics must be upgraded to accomplish submicron beam diameters in production.
- Cathode development must occur to accomplish increased beam brightness with decreased beam size.

If sufficient progress is made in these areas, electron beam memories should be commercially available in 1980 with capacities to 2×10^9 bits for two-stage deflection units with access times of 3.3 μsec (no operation changes involved) and costs on the order of 0.01 to 0.02 ¢/bit. By 1985, similar units should have capacities to 8×10^9 bits, with access times of 2.5 μsec and costs of 0.001 to 0.005 ¢/bit.

Because of the rapid advances being made in semiconductor memory technology and because of the limited research currently going into electron-beam memory technology, electron-beam technology may never reach its projected potential as a viable candidate to fill the memory access time gap between semiconductor and moving magnetic media.

Fixed-head disk storage technology is still evolving and viable although moving-head disks (see below) are evolving faster. By 1985, areal density, spindle capacity, and system capacity will improve 150%, and cost per bit will improve by 30% (drive) to 60% (system).

Moving-head disk storage technology is also still evolving and viable. By 1980, spindle capacities, maximum system capacities, and cost per bit will improve by 50%, and transfer rates will improve by 25% over current values. By 1985, spindle capacities and maximum system capacities will improve by 150%, cost per bit will improve by 80%, and transfer rates will improve by 275% over current values. However, access time will not experience major advances. Fixed-disk packs will become common once again to achieve the tighter mechanical tolerances needed to accomplish higher densities at the least cost.

Floppy disks are currently used in low-cost key entry and low storage capacity applications. Trends in floppy disk technology are toward increasing the storage capacity and transfer rates by increasing

the track and in-track densities, improving reliability by using better media and encoding, and lower power consumption. By 1985, capacity should improve by 300%, transfer rates by 75%, and power consumption and reliability (MTBF) by about 20%.

Cartridge disks are gap fillers between floppy disks and full-sized disks. Some trends in cartridge disk technology are toward more cartridges per drive, more disks per cartridge, and higher areal density. These will cause the storage capacity per drive (and cartridge) as well as the transfer rate to increase and cost per bit to decrease. By 1985, drive capacity should increase by 100%, system cost should improve by 50%, and transfer rates should improve by 35%.

Half-inch reel-to-reel tape, in 7- and 9-track formats, is the major medium used for storage and interchange of digital data throughout the world. Major trends in half-inch tape are toward higher in-track density, higher transfer rates and capacities, and lower cost per bit. By 1985, these measures will have improved 100% for both low-cost and high-performance tape drives, with two exceptions. Cost per bit for low-cost half-inch tape drives will have improved only 50%, while transfer rates for the same class will improve 400%.

Cassette/cartridge tapes are used in data entry terminal and small business systems. Trends include increasing tape speed and more tracks to reduce access time. By 1985, these characteristics will have improved around 100%.

Trends in data processing communications elements are toward use of single microelectronic devices to do all, or important portions of, various communications tasks. This reduces the power consumption (where lower levels of integration were used before) and/or frees the main processor element for other work (where the processor did the communications task before). By 1985, fairly complex byte-oriented I/O channels will be implemented on one small printed circuit board with the main tasks being done by less than 10 integrated circuits. Modems capable of 9,600-baud transmission/reception will be built in the same way, as will other important communications controllers. Multiplexers/demultiplexers with Gbit/sec data rates will also be available.

Read-only memories, particularly MOS and bipolar ROMs and MOS PROMs, are following the same trends as read-write memories. Major improvements will occur in all performance and capacity categories, though ROMs will probably be speed-optimized rather than power- or capacity-optimized. By 1985, chip densities will improve by a factor of 16, access and cycle times will improve by 75%.

DATA BASE SYSTEMS ELEMENTS

Data Base Management System (DBMS) software and/or techniques are receiving wide attention for managing the large data bases of the future. Some of the key issues associated with DBMSs are data independence (loosely defined as the immunity of applications programs to the data base structure), data base security, data models, and data base machines.

Data independence is gradually becoming a reality with certain limitations. It is expected that existing applications programs will essentially be immune to data item changes, as well as to changes that add a relationship, by no later than 1982. At the same time, changing relationships will still be a problem under many circumstances for the foreseeable future.

Present DBMSs provide security at the data base and file level. Projections call for security at the record level to be widely available by 1980, security at the data aggregate level to be standard by 1983, and the individual data item to be protected by 1984. These features will be implemented primarily in software prior to the mid-1980's, when such features will be available in hardware, probably as an option.

Data models are receiving the attention of a number of software and hardware vendors. Current DBMS software is primarily oriented toward hierarchical models. The current trend is away from hierarchical data models toward relational and network models. Although there is very high interest in relational models, the network model is expected to be the primary data model available and in use through 1985. In this respect, a high-level language capable of translating a data base from a hierarchical model to a network model is expected to be available by 1983.

Many authorities are of the opinion that the major mainframe manufacturers will orient their next generation of computers toward data base management, and thus create the so-called data base machine. Such a machine will exhibit a hierarchical memory structure, sophisticated data clustering capabilities, and a full complement of subsidiary DMBS processors capable of performing functions such as searching, sorting, security checking, etc. Projections are that a machine with capabilities such as these is expected to be available by 1982.

INFORMATION DISTRIBUTION ELEMENTS

Terrestrial networks will continue to provide the dominant method of interconnection for data communications in the early to mid-1980's, although long-haul links will tend to use more and more satellite channels. The average user, however, will interface to the satellite via a telephone line in lieu of having a satellite terminal. At the same time, the use of small satellite terminals, particularly for reception of wideband data, will grow at a rapid rate.

Terrestrial networks will still be predominantly analog, and thus require modems, into the mid-1980's, with a continued strong trend toward an all-digital network. By 1985, the major long-haul networks will probably be digital, with a continued dominance of analog circuits to subscribers. Optical fibers will be used extensively in future networks.

Data rates over voice-grade lines are expected to increase to 9,600 bits/sec on multi-point lines, including switched voice-grade lines, and to approach 14 to 15 Kbits/sec on point-to-point lines. The improvements will be primarily the result of improvements in software techniques and modem hardware. However, the use of digital links between switches will also play an important role in increasing the data rates.

Packet-switched networks will continue to increase and will play a major role in computer communications of the future. Small satellite terminals will see increased use (particularly receiving terminals) and will likely prove to be feasible for distribution of wideband NASA data that require fast turnaround.

INFORMATION PRESENTATION ELEMENTS

Information presentation elements include both the dynamic or real-time systems (terminals, graphic display systems, and image presentation systems) and hardcopy systems (COM, printer/plotters, etc.) for presenting data in a viewable form. Advances in capability in all areas of information presentation will result from research in the specific technology areas as well as research in related areas of data processing and memory systems.

The capability, flexibility, and throughput of all types of information presentation systems will increase as the individual devices become more intelligent and thus more independent of the host processor. Flexibility of both dynamic and hardcopy systems will increase as a result of internal processing and memory capacity. In general, prices are likely either to remain constant or to increase as a result of the added capability.

The cathode ray tube will remain the dominant display medium for most applications where size is not a limiting factor. (The size of CRTs is not expected to exceed 25 in.). Plasma panels will become more economical and will be used for applications requiring a display with mechanical ruggedness and a thin profile. Electroluminescent displays will become available on a limited basis by 1980 and will be capable of good-quality video by 1985. The liquid crystal panel displays will become a major competitor of the CRT for devices not displaying motion.

In the hardcopy area, there will be a continued improvement in the area of COM technology, primarily as a result of the use of front-end processors. COM prices will not decline in the foreseeable future, since any reduction in component prices will be reflected in greater sophistication and flexibility.

With few exceptions, printers will employ well-developed and time-proven technology. The low-speed printers will exhibit little to no improvement in speed and will decline only slightly in cost. The higher-speed impact printers will remain constant in capability and will be

replaced by the nonimpact technologies. The very-high-rate (above 20K lpm) nonimpact printers will remain constant in speed but will increase in reliability and decrease in cost.

SOFTWARE

Systems software, particularly operating systems, are an important extension of the hardware, serving to interface applications programs with the hardware while making the hardware configuration invisible to the applications software. Operating systems will continue to keep pace with the hardware in the 1980's. They are expected to be more usable, flexible, and dependable. Hardware technology will aid this development during the 1980's as larger and larger portions of the operating systems will be implemented in firmware.

Software includes both systems software and applications software. Due to the rapid advances of data processing hardware technology in the 1980's, both systems and applications software will become larger (in terms of number of instructions) and more complex.

The three major concerns about software are cost, programmer productivity, and quality. These concerns point to a general concern as to whether the technology that produces the software can continue to meet the demands for the larger and ever more complex programs in the 1980's. Analysis of two aspects of software engineering technology -- language and methodology -- indicates that by 1985 software technology may well become the limiting factor in data systems development. Programming has traditionally been practiced as a craft, and although there is a movement by the industry toward more scientific and disciplined approaches, software engineering is still in its infancy. In fact, it is at least 10 years behind the related discipline of digital engineering (which produces the hardware). It is possible for software engineering to catch up with the demands placed on it by the development of a universal engineering language and increased use of software libraries. However, the industry remains content to invent general-purpose coding

languages and tighter programming controls with the hope that these actions will provide a major "breakthrough". At present, there are no trends to indicate that either the existing solution attempts can provide the breakthrough or that before 1985 the industry will give up trying to make them work. Similarly, there is no trend to indicate that programmer productivity will increase beyond its present 3% growth rate.

CONCLUDING REMARKS

In summary, all areas of data systems technology are advancing at unprecedented rates. Although technology is advancing rapidly and there is a high probability of satisfying NASA's future requirements, technological capability does not imply availability, and developments at the system level may be required in a number of areas. In view of this, a continuing assessment of technology at the component and the system level is required to establish where such developments are needed and how to best accomplish the desired result.

INTRODUCTION

OVERVIEW

Future NASA Office of Applications (OA) programs require the collection and subsequent dissemination of unprecedented volumes of data on a timely basis over widespread geographic areas. Users of the data range from large Government departments (Interior, Agriculture, etc.) to the individual (college professor, researcher, etc.) with a common limitation: budget.

The OA foresaw the technical, logistical, and integration problems associated with such a complex undertaking and implemented the Data Management Program to review and input to those functions that affect NASA's participation in the objectives, including requirements, capabilities, integration, and budget assessment. This report covers one aspect of the capabilities function: i.e., technology assessment. Also, because technology assessment covers cost as well as performance, inputs to budget will result.

The scope of a technology assessment for end-to-end data systems is so massive that it was impossible to provide the breadth and depth of coverage that is desirable within a limited resources program. Selected areas of technology that were considered as critical to NASA's future programs have been given the broadest and most in-depth coverage.

Technology assessment as defined herein is directed toward establishing the state of the art, trends, and future capabilities (1980-1985) for end-to-end data systems. Data systems elements making up end-to-end data systems were defined by the NASA Data Systems Laboratory and are as follows:

- Data Generation - Sensing of physical parameters that characterize the phenomena to be investigated and translation of these parameters into electrical signals acceptable to a data system.

- Data Processing (Space) - Processing of data, usually but not exclusively in the digital mode, to change them to a form more acceptable for data transmission, storage, or use onboard

- Data Storage (Space) - Storage of data for delayed transmission; for delayed processing, display, or other use onboard; or for physical return of the storage media

- Data Handling (Space) - The acquisition and/or distribution of data for data processing, display, storage, or other use onboard. The combining of data streams from several sources or the modification of single data streams to a format suitable for storage and/or transmission.

- Communication (Space to Ground) - All functional elements necessary to transfer data to or from a spacecraft to the ground. Includes the functions performed by modulator, transmitter, airborne antennas, relay systems, ground antennas, receivers, and demodulators

- Preprocessing Elements (Ground) - The data processing elements (hardware and software) that take either predetection or postdetection data and detect (as necessary), demultiplex, and reformat the data into a parallel, digital format with calibration

- Processing Elements (Ground) - The hardware and software components that take parallel formatted digital data (either raw or semiprocessed) and produce end products suitable for either further processing, display, or storage within a data base

- Data Base System Elements (Ground) - The hardware and software components that enable the creation of an integrated data base from logical files and the subsequent retrieval of information from the data base using either specified keys and/or relationships

- Information Distribution Elements (Ground) - The manual and electronic means (hardware, software, and protocol) for requesting and distributing processed and preprocessed data in response to a user request

- Information Presentation Elements (Ground) - The hardware and software elements necessary to provide hardcopy and dynamic visual presentations of image, alphanumeric, and graphic data.

In the following sections, current and future data system technologies are discussed within the context of the above definitions with the following exceptions. Data generator elements, which are covered by a separate contractor study, are not covered in this report. This section, which would have been Section 1, was replaced by a new section entitled "Technology Perspective" that discusses the microelectronic revolution as it relates to the various technologies discussed throughout the remaining sections of the report. Another modification to the report format was the inclusion of the software technology discussion as a separate section (Section 11) rather than as part of ground-based processor technology as orginally outlined. Thus the report is organized into 11 sections with Sections 2 through 10 consistent with the corresponding data system elements as defined above. The elements are discussed in terms of the subelements or levels that make up the parent element. For example, data processing elements make up hardware (processors, data storage, etc.) and firmware (memory and microcode). The subelements are discussed in terms of state of the art (performance capabilities of typically top-of-the-line or high-performance devices), trends (uses, effects on cost and performance, historical extrapolations, etc.), and projected developments (forecasted capabilities for some future time period -- 1980 to 1985 for this report).

PUSPOSE AND USE OF REPORT

The report is intended as the beginning of a data bank that can be used by designers, analysts, and planners of NASA data systems. The trends and projections should enable designers for future NASA data systems to assess particular aspects of the technology such as the various system component technologies to determine reasonable data system design configurations. An individual data component or subcomponent

will frequently have application in several data systems element areas. For example, semiconductor memories could conceivably be used in any of the 10 data systems areas. To minimize repetition of data within the report, discussions relating to a specific data system element or component are presented within the parent element section where its application is expected to most heavily affect NASA data systems. In the above example of semiconductor memories, the subject is covered under Data Processing Systems Elements (Section 7) and is referenced in other sections where applicable.

Finally, an important consideration concerning the technology forecasts is that a forecast does not necessarily imply availability. Further, a list of specifications does not imply that all specifications within a given list will be feasible within the same system. For example, speed and power consumption are usually opposing considerations (i.e., as the propagation delay within a circuit decreases, the power increases) and the specification for low power consumption per bit may not correspond to the best projections for access and cycle times. To clarify limiting factors such as these, verbal discussions that attempt to identify pertinent considerations are presented within the report. Thus the user must interpret data herein within the intended context for it to be meaningful.

APPROACH

The approach to developing the assessment was based on a combination of analysis of literature in the different technology areas and a sampling of opinions of experts and authorities currently working the areas.

Many authorities in the data systems field responded to requests for information via the mail, telecons, and visits, and significant parts of the forecasts are the result of their contributions. Many of the respondents generously gave permission to use their names as the source of the information. However, the sponsoring agency (NASA) requested that specific references be made only to published literature

to eliminate the possibility of embarrassment to either the individual or the organization. Individuals and organizations that contributed are listed alphabetically at the end of the report. Many of these individuals are outstanding in their fields. Their contributions are gratefully acknowledged.

As with literature, the responses of individuals regarding future capabilities varied rather widely. These responses had to be weighted based on some criteria. Frequently, the criteria used were in favor of the majority of responses. Another criterion was the availability of prototype hardware that appears to have potential as production equipment in the timeframe of interest. Finally, the past experience of the individual within a given area was weighted heavily. For example, if an individual or his organization has been dominant in a given field for a reasonable length of time, then the opinions of that individual were weighted heavily.

SPECIAL CONSIDERATIONS

A number of traps exist in any report of this type, even when reporting state of the art. One of the first traps is to distinguish in the literature and in replies to questions between what is feasible on a production basis versus what is available on a custom-design (higher-cost) basis. Beyond this, one must distinguish between what is available on a custom-design basis and what is available only in the laboratory; and finally, for space applications, there must be a distinction between what is available for terrestrial applications versus space applications. The literature is usually vague on these distinctions.

Another very important consideration is that just because technology will support a capability, this does not mean that the capability will be available. The availability in many instances is a function of market demand or emphasis in the area of concern. Microprocessors, semiconductor memories, and other integrated circuit devices, for example, provide the capability for building tremendous flexibilities

into data system elements, but the capability does not imply availability. Similarly, various elements of data systems comprise several underlying technologies: i.e., RF technology and integrated circuits. Thus developments in one technology -- e.g., IC speed -- that would permit an advance in a data system element capability is not necessarily reflected on a one-for-one basis because of either the limitations or lack of emphasis in another area. Attempts are made to account for this in the discussions where possible.

In summary, an attempt was made to distinguish between what is likely to be available commercially and what is feasible in terms of technology. Projections are based on the use of commercially available components; thus developments at the component level will not be required. Developments at the system level (above the IC level) may be required.

REPORT FORMAT

The report is formatted to comply with NASA's data bank organization and to facilitate update. As a result, a large number of pages are only partially filled.

1. TECHNOLOGY PERSPECTIVE

During the last decade, the evolution of electronic technology has been so dramatic that it has often been referred to as a revolution. The common denominator in this revolution, whether it be space data systems or digital watches, is the "miracle" of the chip. Although there have been significant advances in a number of other technology areas that relate to NASA end-to-end data systems requirements, none can compare with the advances in microcircuit technology. The complexity of microelectronic devices has increased exponentially during the last 15 years, and all indications are that the microelectronics revolution is far from having run its course. In fact, it is currently projected that the revolution should continue at its present pace for at least another 10 years or more. It is for these reasons that the technology perspective presented in this section focuses on the microelectronics revolution and attempts to project where it is headed within the 1980-1985 timeframe. The information presented in this section thus provides a common thread for many of the projections found throughout the remaining sections of this report.

Another important aspect of the technology perspective presented in this section is that it emphasizes the need for accurate technological forecasting. Now, more than ever before, timing is most critical in the development and/or procurement of large data systems or data system elements. One example that underscores the tremendous current economic pressures resulting from advancing technology is the recent price reduction of large-scale computers by most of the industry, including the largest price reductions ever made by IBM. These price reductions are representative of the trend in improving cost/performance that is expected to continue for the next decade, as explained in Section 7. This example also illustrates the importance of using technological forecasting as a tool to reduce the cost of data systems development and/or procurement.

As a result of the microelectronics revolution, data system technology has advanced to the point at which requirements are beginning to assume their rightful role as driver of the data system design. This will increase at an accelerated rate in the 1980-1985 timeframe. Future space data systems will have access to new technologies that are capable of enhancing every aspect of space data management from the space sensor through the final processed and displayed data on the ground. The designers of these data systems will have considerable flexibility in where the data are processed, where and in what form they are stored, when and how the data are transmitted to the ground, and how they are handled on the ground. Capabilities available for future data systems include:

- Data management systems with adaptive features and interactive capabilities (including limited voice interaction) onboard and via space-to-ground communication links
- Space processing options that range from sophisticated dedicated processing to shared resource processing, with on-line, fast-access storage for data and complex software algorithms. Processing capabilities may be distributed at any point from the sensor to the final display.
- High-capacity data base storage (10^{11} to 10^{12} bits) onboard the spacecraft with update capabilities either onboard or via the ground-to-space communications link
- Programmable/adaptive data handling systems with asynchronous, variable-bandwidth, priority-controlled message handling, using store and forward techniques
- Communications power and bandwidth capabilities to satisfy the most complex orbital requirements
- More powerful ground data processing and data management capabilities through faster processors and memories, more sophisticated system architectures, and more sophisticated vendor software
- Capabilities for distribution of wideband data via small satellite terminals, and increased terrestrial capabilities employing packet-switching techniques and advanced hardware technologies.

The ability to implement the preceding capabilities will be based to a large extent on the availability of microelectronic (integrated circuit) devices, including extremely low-cost computers and high-capacity memory devices on a chip. The need to share computer and memory resources onboard the spacecraft will be nil. Each subsystem will have its own processor and storage devices, except in instances where the same data are shared by many users. Even in the case of shared data, there will be instances where it is cost/performance-effective to duplicate data in different memories onboard a spacecraft. Thus distributed processing and distributed data bases onboard the spacecraft will be the norm.

This section addresses the technologies that will be available to support future end-to-end system designs from the standpoint of current trends and projected developments in basic microelectronic technology. The capabilities of specific data system elements at the subsystem level are discussed in the remaining sections of the report.

1.1 MICROELECTRONIC TECHNOLOGY

Microelectronic devices, also designated Integrated Circuits (ICs), play an extremely important role in present-day data systems, and this role is projected to increase significantly in future data system designs. This section discusses the technology of microelectronic devices in general. Other sections of the report discuss specifics of microelectronics as they relate to microprocessors, memory devices, peripheral devices, communications, display systems, and related subjects. Discussion in this section centers on factors that are common to all microelectronic devices, such as state of the art in microelectronic technology, how microelectronic devices are used in current data systems, what factors are affecting future development of microelectronic devices, and future projections for microelectronics. The bibliography lists a number of excellent articles that provide a more detailed discussion for those readers who wish to pursue the subject in greater depth.

1.1.1 Background

The history of microelectronics began with the development of the transistor by W. Shockley of Bell Laboratories in 1947. During the 1950's the digital computer provided a large market for the transistor and stimulated research into methods of further microminiaturization. Most of the initial research was aimed at miniaturization of conventional computers and/or hybrid packaging techniques. These techniques were in general expensive but in many cases were adapted for aerospace applications where the additional cost could be offset by significant reductions in weight.

The origin of modern integrated circuit technology can be traced to the development of the planar process at Fairchild Semiconductor in 1959. The planar process was based on two important process patents. The first, by K. Lehovec of Sprague Electric, provided a means for electrically isolating the individual transistors from the integrated circuit substrate, thus allowing a circuit to be fabricated on a single substrate with multiple active devices. The second invention, by

J. Hoerni of Fairchild, provided a method for using silicon dioxide as an insulator between the interconnecting metalization patterns and the underlying semiconductors.

In 1964, five years after the production of the first planar transistor, Gordon E. Moore observed that the number of components in advanced integrated circuits had been doubling every year....Mr. Moore, who was then the director of research at Fairchild, predicted that the doubling process would continue into the foreseeable future. Today, after 19 years, there has been no significant deviation from Moore's original projection (also called Moore's Law). Moore's Law, illustrated in Figure 1.1.1-1, extrapolates the forecast to 1984 and overlays some of the circuit devices that have kept pace with the law. If this trend continues, the number of components per chip for production integrated circuits will exceed 30 million by 1984. The probability of this trend continuing is discussed in the following subsections.

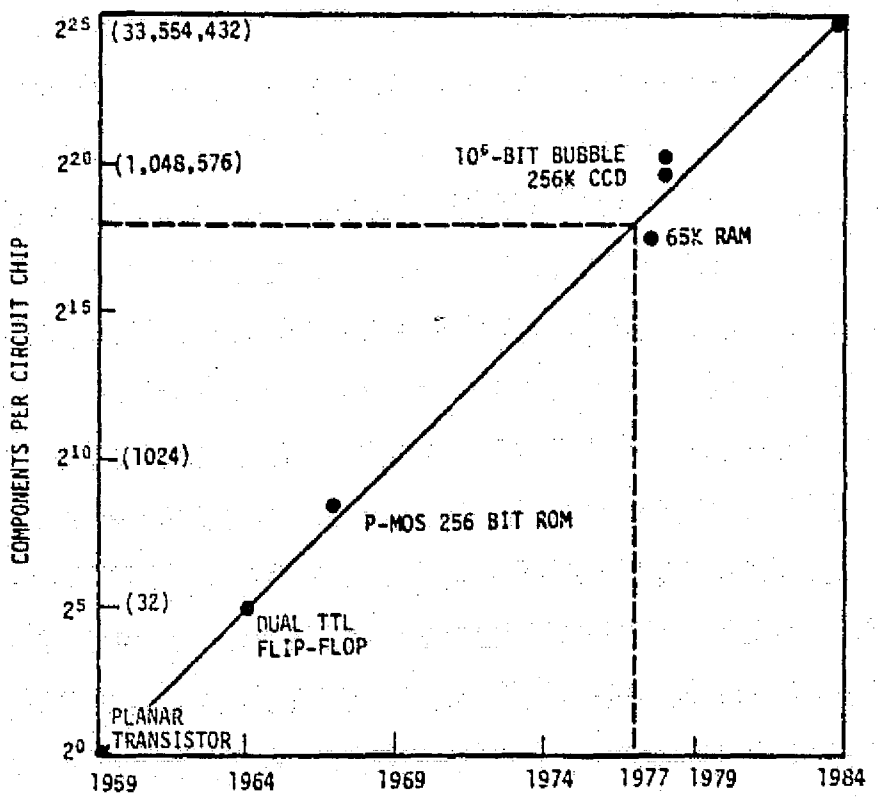


FIGURE 1.1.1-1. MOORE'S LAW ILLUSTRATED FOR 1959-1984

1.1.2 State of the Art in Microelectronics

Electronic technology has progressed through several levels of integration [small-scale integration (SSI), medium-scale integration (MSI), and large-scale integration (LSI)] to the point where very-large-scale integration (VLSI) is beginning to emerge, with initial production devices providing complexities in excess of 100,000 circuit elements (resistors, capacitors, transistors) per chip. Laboratory devices that contain more than one million circuit elements have been developed, and techniques capable of dwarfing these figures are being researched.

The state of the art in terms of production devices that are currently being built on a single chip includes a variety of digital and analog devices, as well as mixed devices capable of handling both analog and digital signals. Some examples of state-of-the-art single-chip microelectronic devices are: 16-bit microcomputers with 2,048 bytes of ROM and 144 bytes of RAM on the chip, 65-Kbit random access memories, 262-Kbit sequential storage chips, and complete multi-channel data acquisition systems that include a 16-channel analog multiplexer and a 10-bit A/D converter.

Advances in microelectronic devices are normally discussed in terms of digital devices. The major emphasis in this report is also on digital devices since the trends for future data systems are so heavily oriented in that direction. Actually, however, significant advances are occurring in both analog and digital devices. Analog devices that are available as monolithic (one-chip) devices include operational amplifiers, power amplifiers, isolation amplifiers, filters, regulators, oscillators, mixers, analog switches, multiplexers, and other related components. Beyond the fact that these components can be fabricated as monolithic units, mixed processing techniques enable the development of linear devices with very unique characteristics. As an example, amplifiers can be fabricated with FET inputs and bipolar outputs. The FET input operates with low input bias current as a result of the high input impedence, but the bipolar output provides good linearity on the output side. Both of these advantages are not achievable with either an all-FET device or an all-bipolar device.

Simultaneously, the technology permits mixing of analog and digital circuitry on the same chip, including but not limited to analog-to-digital and digital-to-analog converters. One unique device in this category that is nearing market availability is a programmable CCD filter. The device operates as a transversal filter under the control of a separate microprocessor and read-only memory. The device can be factory programmed to operate as one of four types of filters, low-pass or bandpass, each in narrowband or broadband versions. This device is discussed further in Section 2.2.

Because of the complex nature of current microelectronic devices, the large number of integrated circuit device processes currently in use, and the large number of microelectronic manufacturers, the exact state of the art of microelectronic technology is difficult to assess except in terms of the general complexity and performance characteristics of available devices and the fundamental limitations of current device fabrication techniques. Since performance characteristics for several state-of-the-art devices are discussed in subsequent sections, only the latter topic is discussed here.

For those readers who are interested in greater detail on how state-of-the-art large-scale integrated circuit devices are manufactured, an article from the September 1977 issue of Scientific American is suggested for further reading. The article, "The Fabrication of Microelectronic Circuits" by William G. Oldham of the University of California at Berkeley, is very readable and contains a number of useful illustrations that demonstrate fabrication and packaging processes. The article discusses the factors that currently limit the economical production of high-complexity microelectronic devices such as the effects of circuit size on process yield. The article also discusses several advanced wafer processing techniques, including X-ray lithography and electron-beam lithography. Additional information and future predictions for semiconductor fabrication techniques are discussed in Subsection 1.1.3.

1.1.3 Trends and Projected Development in Microcircuit Complexity

This subsection discusses the outlook for the continued evolution of microcircuit device complexity. At the present time, semiconductor* microelectronic devices are beginning to emerge with complexities in excess of 20,000 circuit elements for logic devices (Ref. 1-1) and in excess of 130,000 circuit elements for memory devices (Ref. 1-2). Measured in terms of logical functions and/or bits of storage, these chips represent complexities of approximately 2,000 gates with 11,000 bits of read-only storage and 65,000 bits of random-access storage, respectively. As complex as these devices are, they are not nearly as complex as the devices projected for 1990 by Dr. Carl Hammer, Director of Computer Sciences, Sperry Univac (see Table 1.1.3-1). In a recent paper (Ref. 1-3) Dr. Hammer predicted that by 1990 memory devices would contain up to 800,000,000 bits per chip and logic arrays (processors) would contain up to 6,000,000 logical functions per chip. Measured in terms of today's technology, these projections represent an increase in chip complexity of at least a factor of 2,000 (or 2^{11}) over today's technology. As pointed out by Dr. Hammer, even these 1990 projections are still well within the macroscopic domain; i.e., still not approaching atomic dimensions.

TABLE 1.1.3-1. COMPARISON OF STATE-OF-THE-ART LSI DEVICE COMPLEXITY WITH DR. CARL HAMMER'S 1990 PROJECTIONS

	MEMORY CHIP COMPLEXITY	LOGIC CHIP COMPLEXITY
1978	1.3×10^5 transistors	2.2×10^4 transistors
1990	8×10^8 transistors	6×10^6 transistors
Projected Improvement Factor	6.1×10^3	2.7×10^2

*Nonsemiconductor microelectronic devices, such as bubble memories, of even greater complexity are scheduled for commercial production by mid-1978.

In order to understand how the complexity of future microelectronic devices may evolve, it is appropriate to first discuss the relationships between minimum device fabrication geometries and logical/storage cell dimensions. These dimensions can then be related to the practical and ultimate limits of the various process technologies to provide future device complexity projections.

If we call the minimum circuit feature W, then the area of a memory storage cell or circuit device (transistor, diode, etc.) can be expressed as NW^2. Two other measurement units commonly used to express device size or memory cell size are square mils and square micrometers. The area of the device can be expressed either as the number of circuit functions per square unit of measure or as the number of square units of measure per circuit function. Since one mil is equal to 25.38 μm, one square mil is equal to 644 μm^2.

The area of state-of-the-art memory storage cells varies from 4 W^2 for contiguous disk bubble memories (see Section 3.2) up to 52.2 W^2 for dynamic MOS random-access memories (Ref. 1-2). Static and other less efficient memory configurations have storage cells that are larger than 100 W^2. Since complex logic devices often contain a mixture of random logic and structured storage arrays, a similar measure of average device size is less meaningful. However, state-of-the-art microprocessors have device densities as low as 60 W^2 (Ref. 1-1).

Currently, most state-of-the-art production LSI circuit devices have minimum circuit features of from 4 to 6 μm. The optical lithography systems used by most manufacturers employ high-pressure mercury arc lamps with ultraviolet output in the 200- to 600-μm range. Depending on the lithography system used, the minimum circuit features that can be printed routinely vary from about 4 μm using proximity printing with a 10-μm gap down to 1.6 μm for f-2.55 projection printers. Some of the 65K memory circuit devices scheduled for production in late 1978 or early 1979 are expected to employ 2-μm circuit features. Although 1 μm was previously considered an ultimate limit for optical lithography,

several companies are currently researching device features down to 0.5 μm using deep ultraviolet lithography.

X-ray lithography and electron-beam lithography are currently the two most viable techniques to eventually replace optical lithography and thus overcome the limitations of resolution and depth of focus imposed by the wavelength of light. Electron-beam systems have been touted as the next logical step in the fabrication of microcircuits. These systems are currently employed on a limited basis to produce masks that are then used in the standard lithographic processes. Further, some E-beam systems are used as direct-wafer-exposure devices for laboratory- and/or custom-developed chips. The general feeling among American authorities is that the initial cost, the low throughput, and the yield forbid the use of these systems for direct wafer production for at least 3 to 4 years. On the other hand, the Japanese have based their entire VLSI program on the use of E-beam fabrication devices. Some of the advantages of scanning electron-beam lithography systems are:

- Computer control rather than masks is used to form the pattern and provide a high degree of pattern flexibility.
- Microscopic pattern distortions can be compensated for in real time.
- They provide excellent pattern registration and overlay.

The major disadvantage of scanning electron-beam systems has been the difficulty in achieving the necessary throughput for cost-effective use.

IBM has been researching electron-beam fabrication systems for the past 15 years and currently has a system (designated EL-1) which has a throughput rate of 22 wafer exposures per hour based on 2.25-in. wafers with 2.5-μm minimum features. The system design is such that it may be adjusted to make exposures in the 1-μm range.

Table 1.1.3-2, extracted from an article in the June 7, 1977, issue of Electronics by L. Altman and C. Cohen, entitled "The Gathering Wave of Japanese Technology", shows the goals of the Japanese VLSI

program using direct-wafer-exposure E-beam processes. The program calls for production devices with minimum line widths of 0.02 mils (0.5 μm) by 1983. In order to achieve 30 million components per chip, as discussed in Subsection 1.1.1, the circuit dimensions must be no larger than 0.2 μm. Although the Japanese VLSI program will most likely not reach 0.2 μm by 1984, it is probable that the competition for business in this highly profitable area will push American firms to achieve this level of complexity by that time.

TABLE 1.1.3-2. GOALS OF JAPANESE VLSI PROGRAM

	RANDOM-ACCESS MEMORY		MICROPROCESSOR		MINIMUM LINE WIDTH (mils)
	CHIP SIZE (× 1,000 mil²)	BIT DENSITY (Kbits)	CHIP SIZE (× 1,000 mil²)	WORD LENGTH (bits)	
1976	32	16	52	8 and 16	0.2
1979-1980	32	64	50	16	0.07
	50	256	60	32	0.04
1981-1983	45	256	55	32	0.03
	60	1,024	65	32	0.02

If 0.2-μm line widths are not achieved by 1984, it is only a matter of time until line widths of this size are capable of being fabricated. Another fabrication technique that shows great promise for achieving line widths down to 0.1 μm and less is X-ray lithography. In essence, X-ray lithography is similar to optical lithography except that soft X-rays are used to expose the semiconductor material instead of ultraviolet radiation. Masks for this process can be generated with standard electron-beam devices. A number of problems remain to be overcome before X-ray lithography becomes feasible. Among these problems are sufficiently sensitive resist materials, improved alignment and pattern registration accuracies, and dimensional stabilities of substrate materials. The December 1977 issue of IEEE Spectrum reports that MIT's

Lincoln Laboratory has used an optical interferometric technique to achieve 0.02-μm alignments. The technique is said to be compatible with both photo and X-ray lithographic systems and is capable of a superposition precision on the order of 0.01 μm. Other researchers at Lincoln Laboratory reported the development of very-low-distortion polyamide membrane masks that are used in X-ray lithography.

In addition to improvements in the area of exposing the lines and in resists, significant improvements are required in the etching process. Up to now, most etching has been done with wet etchants. Wet etchants generally cannot be used for line dimensions below 2 μm. Some manufacturers use dry etchants with a reasonable degree of success, but the general belief is that plasma etching offers the best promise for the future, with known capabilities well below 0.5 μm.

Authorities generally agree that electron beam direct-wafer-fabrication devices are 3 to 4 years away, although as pointed out earlier, Japan plans to use E-beam direct fabrication techniques as production devices in the 1981-1983 time period or earlier. Research in the area of electron beam devices is continuing, including the development of a variable-shape electron beam that substantially reduces the number of exposures required to create a circuit pattern. Similarly, X-ray lithography is believed to be at least 5 or more years away at this time. Probably all authorities would agree that line widths of 0.2 to 0.1 μm can be achieved during the 1980's and that the market demand will force the fabrication of chips with these line widths. Whether it will take place on a production basis by 1984 is speculative at best, but the probability of building custom design devices with these circuit features is very high. In fact, laboratory devices have already been built with circuit features even smaller than 0.2 μm. Figure 1.1.3-1 presents projections of minimum circuit features necessary to satisfy Moore's Law and compares it with the Japanese VLSI program, laboratory devices that have been built, and resist resolutions. Points 8 and 9 of this figure are noteworthy since they represent laboratory devices that have been produced by IBM and Hughes with circuit features that are smaller

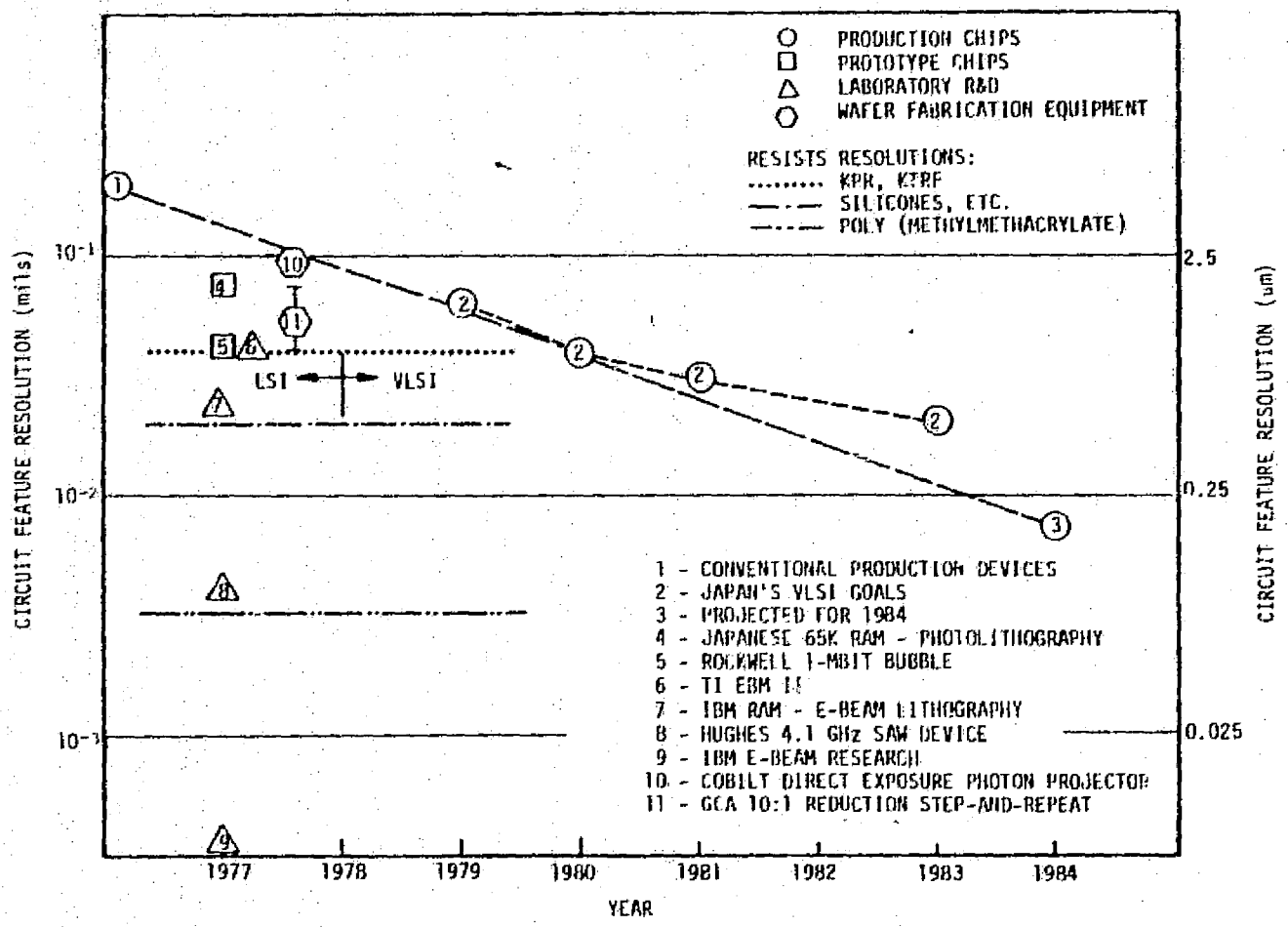


FIGURE 1.1.3-1. PROJECTED EVOLUTION OF MINIMUM INTEGRATED CIRCUIT FEATURES

than the 0.2 μm required to sustain Moore's Law. Both of these devices were produced using E-beam fabrication techniques, which should be rather widespread in the mid-1980's. Simultaneously, the figure shows that optical lithography is continuing to progress and will fill the gap until such time as E-beam technology becomes economically feasible for production environments.

A related technology that cannot be overlooked when considering microelectronic devices is magnetic bubble memories. The Rockwell 1-Mbit bubble memory (point 5 on Figure 1.1.3-1) has 1-μm features and uses the asymmetric disk structure with 1.8-μm bubbles and an 8-μm period to provide a storage density of 1.5×10^6 bits/cm^2 (see Section 3.2). Projecting the minimum circuit feature to 1984, or 0.2 μm, and changing the structure to the asymmetric chevron, which provides twice the density of the asymmetric disk, yields a bubble device with a storage density of 7.5×10^7 bits/cm^2, or a storage capacity of 67 Mbits for a relatively small-geometry chip. Such a device would have in excess of 67 million circuit elements (or propagate elements). If we further assume that the contiguous disk bubble structure is used, which can provide at least four times the density of the asymmetric chevron, and that the size of the chip is increased to 4 cm^2 (which should be reasonable by 1985 using electron-beam fabrication with step-and-proceed and adjacent field stitch capabilities), the storage capacity of the resulting device could exceed 10^9 bits. Of course this assumes that stable bubbles can be scaled down and detected with diameters of 0.05 μm. However, even this circuit density would still only be using device features with areal densities that are 225 times less than might be achieved if the best current research (0.008 μm) ever reaches production status.

Four technical conditions are necessary to achieve superscale integration:

- The dimension of minimum circuit features on a chip must decrease to submicron dimensions.
- Logic technologies with very low power-delay products must be developed.

- Circuit characteristics must scale proportionately to circuit features or new device families must be developed that can be scaled down.

- In relation to circuit characteristics, any changes necessary to achieve essentially proportional characteristics (e.g., voltage scaling) must not degrade system performance and/or reliability within the intended operating environment.

Some of the new solid-state technologies that may satisfy these conditions are discussed in the following section.

1.2 EVOLUTION OF SOLID-STATE TECHNOLOGIES

Evolution in solid-state technology has taken place at many levels, including the device level, the device family level, the degree-of-integration level, and the solid-state material level. This section emphasizes the historical developments in materials and the degree of integration used, the two most slowly evolving levels. Subsequent sections also cover device family and device evolution as well as developments in solid-state technology.

Evolution of semiconductor materials, presented in Table 1.2-1, has been from selenium to germanium to silicon to gallium arsenide. Each of these materials has exhibited one or more properties, such as lower leakage currents or higher carrier mobility, that has made it more desirable than its predecessor for use in semiconductor devices. However, the use of a new material has in general not eliminated use of predecessor materials. Instead, these predecessor materials have continued to be used in special-purpose applications, and even in simple, nondemanding new applications. Future development in semiconductor materials will be in the direction of compound semiconductor materials such as gallium arsenide or indium phosphide. Indium phosphide, as well as other new semiconductor materials, is currently being researched.

Some nonsemiconducting materials are also being researched for use in other new types of solid-state devices. The evolution of these materials is presented in Table 1.2-1. The two most important device types using these materials are magnetic bubbles and Josephson junctions.

Evolution of the level of integration used in solid-state devices, also presented in Table 1.2-1, has been from discrete elements to small scale to medium scale to large scale to very large scale. At each of these levels, more functional devices have been packed into a single "chip". However, the introduction of a new level of integration has never completely eliminated the use of predecessor levels. Devices of a lower level of integration (down to the individual diode or transistor) are generally required to "stick together" the larger-scale devices in even the most modern designs.

TABLE 1.2-1. EVOLUTION OF SOLID-STATE TECHNOLOGY*

MATERIAL	DEVICE TYPE	LEVEL OF INTEGRATION	YEAR
Selenium	Diode	None	1930's
Germanium	Diode	None	1950
	Transistor	None	Early to mid 1950's
Silicon	Diode	None	1954
	Transistor	None	Mid to late 1950's
	Integrated Circuit	Small Scale	Early 1960's
		Medium Scale	Mid 1960's
		Large Scale	1969
		Very Large Scale	Now in late development and early production
Gallium Arsenide	Diode/Transistor	None	Early 1970's
	Integrated Circuit	Medium Scale	Now in late development
		Large Scale	Now in research and early development
		Very Large Scale	Now in research
Garnet** etc.	Magnetic Bubbles	Large to Very Large Scale	1977
Various**	Josephson Junction	Various	Now in early development
Indium Phosphide and others	Integrated Circuits	Various	Now in early to mid research

*Semiconductors were first employed by Heinrich Hertz for detection of wireless radio waves in 1889. Dates given in this table are for commercial application.

**Non-semiconductors

1.2.1 Life Cycle of Solid-State Technologies

Solid-state technologies at any level (individual devices, device families, and materials) have a limited life cycle. This life cycle, illustrated in Figure 1.2.1-1, is characterized by a general rise which is then followed by a fall in interest. The time axis of life-cycle chart can be subdivided into four identifiable though overlapping regions: research, development, production, and phase-out. Figure 1.2.1-1 shows some device families and materials in their relative current positions with respect to these regions and each other.

Figure 1.2.1-1 is not meant to imply that devices, families, and materials traverse the life cycle at the same rate. Individual devices traverse the cycle faster than their respective families, and device families traverse the cycle faster than the material from which they are made. Furthermore, some devices/families/materials complete the cycle faster than others, and the "speed" with which a device, family, or material traverses the cycle may not be constant. For example, a device may stay in research and development for a long time and in production for a short while, or visa versa. In attempting to use the life cycle concept to predict the future, one further complication arises: namely, "false starts". Many device families (and to a lesser extent, individual devices and materials) show early promise but never get beyond the research or early development phase. This often happens because of a single practical problem that cannot be overcome, or because some newer technology with even more promise catches up on the life cycle and causes researchers/developers to lose interest in the older circuit family or material.

1.2.2 New Solid-State Device Technologies

This subsection addresses a number of new solid-state device technologies, with particular attention paid to their potential as high-speed and/or low-power (high-density) logic devices and memories. Figure 1.2.2-1 compares the gate delay, power, and power-delay (or speed-power) product of several advanced technologies. Few of these technologies are far enough along in their life cycle to be considered for current designs,

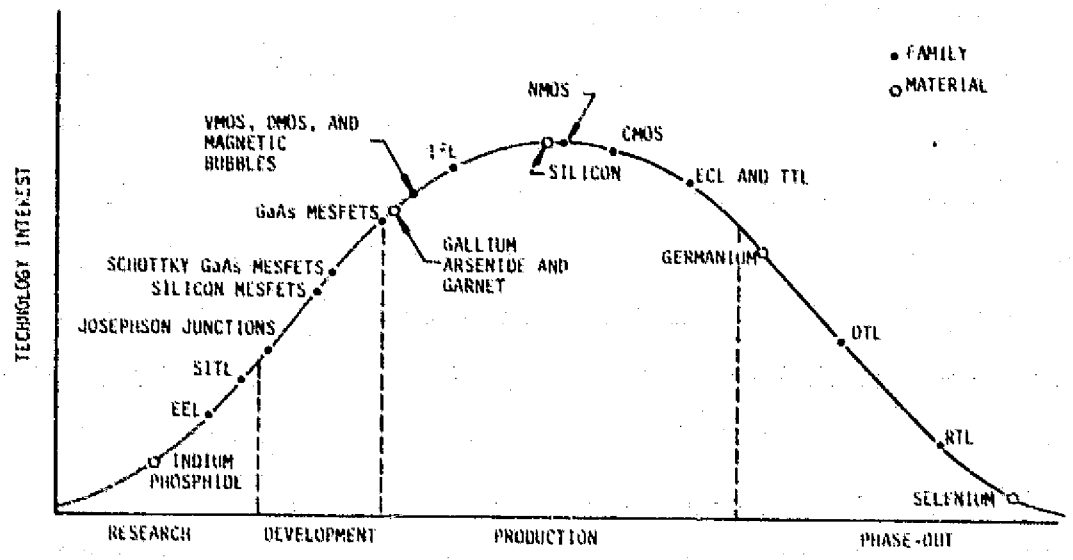


FIGURE 1.2.1-1. LIFE CYCLE OF A TYPICAL SOLID-STATE DEVICE/FAMILY/MATERIAL

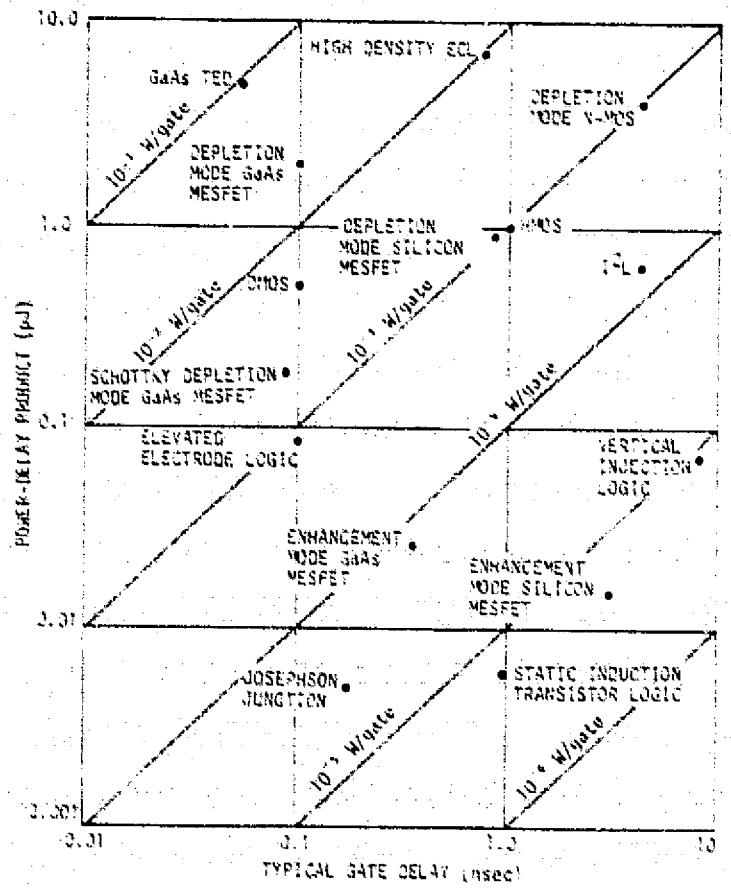


FIGURE 1.2.2-1. COMPARISON OF TYPICAL GATE DELAYS AS A FUNCTION OF THE POWER-DELAY PRODUCT AND POWER PER GATE FOR VARIOUS ADVANCED LOGIC TECHNOLOGIES

and some may never get beyond research or early development (the "false start" problem). Despite this, some of them will be used in the next decade for high-speed and/or high-density (low-power) logic and memory system designs. In general, a device's technology is considered better if it either reduces power, increases speed, or does both simultaneously.

There are several things suggested by Figure 1.2.2-1. One is that some of the newer device technologies are just as fast as or faster than ECL yet have per-gate power consumptions compatible with LSI and VLSI designs. Another is that it appears that some of the newer device technologies appear to be closing the gap between the performance semiconductors and Josephson junctions.

1.2.2.1 New Silicon Technologies - A number of new silicon device technologies have been reported in the last few years. Some of them, notably VMOS and DMOS, are already in the production phase of their life cycle. Others, such as silicon MESFETs, EEL, and SITL, are still in research or development.

VMOS (V-groove MOS), DMOS (double-diffused MOS), and HMOS (high-performance MOS) are variations of NMOS technology. HMOS, in fact, is essentially a scaling down of "traditional" NMOS by using higher-resolution production techniques to produce finer lines and spaces and thus smaller devices. This is a rather straightforward extension of the scaling technique that has been used in integrated circuit technology since its beginning. HMOS will not be considered further.

VMOS is a new MOS technology in which a V-shaped groove is formed at the site on the silicon wafer at which a transistor is to be formed, and the diffusions usually done on the flat surface of the silicon are done on the sides of the groove. This process may save some space but more importantly it shortens the length of the channel that the carriers must traverse. This, of course, speeds up the switching of the transistor and thus of the entire device.

DMOS, also known as DSAMOS (diffusion self-aligned MOS), is another short-channel NMOS technology. In DMOS, the effective channel length is shortened by diffusing two areas, one lightly and the other

heavily doped, where only one heavily doped area is diffused in traditional NMOS. More information on DMOS and VMOS, especially as they apply to memory products, may be found in Subsection 7.2.1-2.

Static Induction Transistor Logic (SITL) is a new silicon technology that uses both MOS and bipolar techniques in a single logic element. Static Induction Transistors (SITs), a type of vertical FET (VFET), are used as output devices, with a regular bipolar PNP transistor used as an injector. SITL shows some promise for VLSI nanosecond to subnanosecond logic.

Elevated Electrode Logic (EEL) is a new silicon bipolar technology that shows some promise for LSI subnanosecond logic. It is currently at too early a stage of development to accurately predict its future.

Silicon MESFETs (metal semiconductor FETs) are currently eclipsed by work on the faster GaAs MESFETs. Nevertheless, Si MESFETs may find some application in the very-low-power (very-high-density) medium-speed area.

1.2.2.2 Gallium Arsenide Solid-State Technologies - The use of gallium arsenide (GaAs) as the basic semiconductor material (in lieu of silicon, which is currently used in practically all integrated circuit devices) for digital logic and integrated circuits has been under investigation in the microwave field for several years. Some properties of GaAs, such as high electron velocity and mobility, make it more suitable than silicon for very-high-speed circuits (e.g., beyond 2 GHz, the current limit of silicon planar logic circuits). Because of this high-speed ability, GaAs is already being used in microwave communications devices as frequency dividers, shift registers, etc. (Subsection 5.1.1.1 provides additional data.)

Digital GaAs devices may be realized as Transferred-Electron Devices (TEDs), Metal-Semiconductor Field-Effect Transistors (MESFETs), or Junction Field-Effect Transistors (JFETs). To date, most accomplishments in the field have been with MESFETs, in depletion-mode (normally on),

enhancement-mode (normally off), and Schottky diode versions, but work is going on for all the device types. Figure 1.2.2-1 illustrates the speed, power dissipation, and speed-power (power-delay) product for these (and other) devices. Note that Schottky diode MESFET logic outperforms depletion-mode MESFET logic in that it can be both faster and lower power. Thus depletion-mode MESFET logic receives little attention here.

Figure 1.2.2.2-1 shows the construction of several types of FETs. Comparing parts a and b of that figure illustrates the difference between a MOSFET and a MESFET. For all FET devices, a voltage at the gate contact regulates the source-to-drain current. For enhancement-mode FETs (Figure 1.2.2.2-1 a, b), no source-to-drain current flows unless the gate voltage goes positive and creates a surplus of electrons in the P-type region near the gate. For depletion-mode FETs (Figure 1.2.2.2-1 c), source-to-drain current flows normally, but is inhibited by a negative gate voltage that causes a lack of electrons in the thin N-type region near the gate. For either type device, a source-to-drain current change may be transformed to a voltage change suitable for driving the gate of another FET by having it flow through a resistive load.

High-speed MESFETs are realized with depletion-mode devices. These may be used as the basic "switches" for logic gates. Because of the depletion-mode characteristics of MESFETs, the voltage change at the load is not compatible in level with the input voltage change, so the voltage level must be adjusted to drive the next stage (see Figure 1.2.2.2-2 a). Actual laboratory devices using depletion-mode GaAs technology MESFETs include basic logic gates, master-slave flip-flops, frequency dividers (using flip-flops), counters, and data multiplexers. Clock rates in the 2-to-4-GHz range for the flip-flops, frequency dividers, etc., have been reported, while propagation delays of 100 psec for the basic gates (corresponding to an equivalent clock rate of about 5 GHz) have been reported. Power dissipations of 20 mW per gate have been reported, thus yielding a speed-power product of 2 pJ. Experiments have outlined methods of fabricating depletion-mode GaAs MESFETs with processing yields of 96% per functional logic gate. Therefore, taking into consideration the normal limits of power dissipation per chip (about

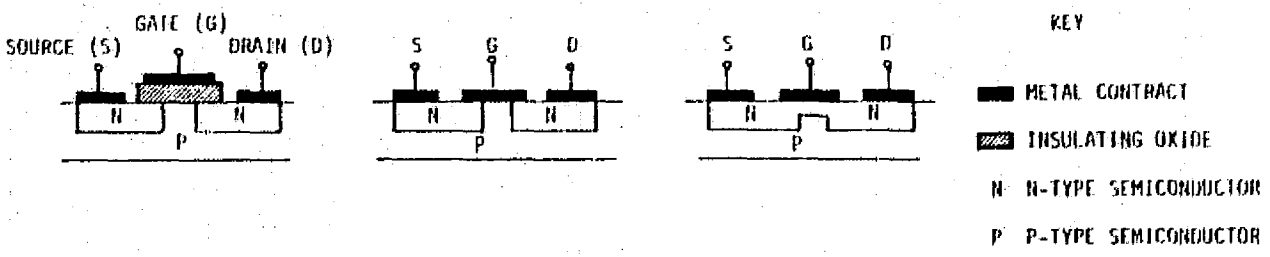


FIGURE 1.2.2.2-1. COMPARISON OF MOSFET AND MESFET CONSTRUCTION

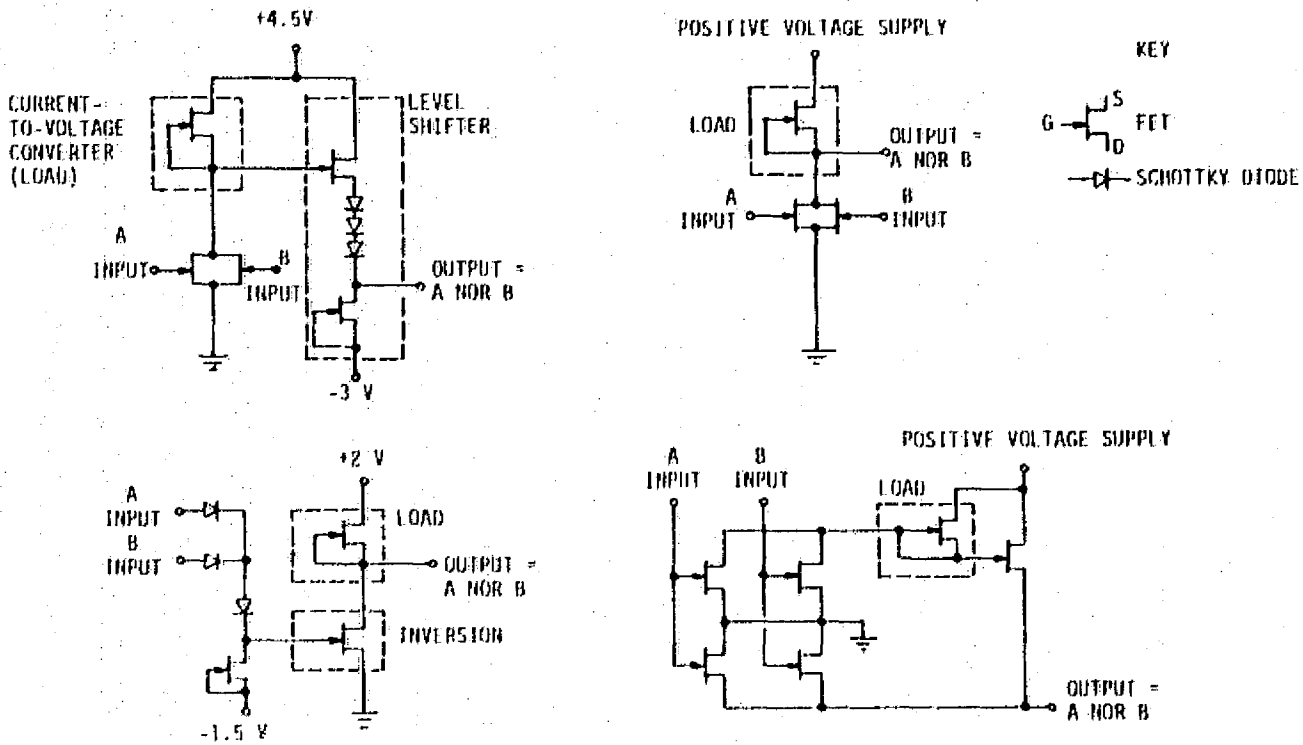


FIGURE 1.2.2.2-2. MESFET LOGIC GATE CONSTRUCTION (TWO-INPUT-NOR)

1 W) and the economic necessity of reasonable fabrication yield per chip, ICs using normally on GaAs MESFETs are currently limited to approximately 50 gates/chip. This limits ICs using this logic to small-scale integration (SSI; i.e., about 10 gates/chip) and medium-scale integration (MSI; i.e., 10 to 100 gates/chip).

Schottky diode depletion-mode MESFET combinations are under investigation for use in high-speed, low-power logic devices. The Schottky diodes are used for most logic functions except inversion and gain, which are attained with depletion-mode Schottky gate GaAs FETs (see Figure 1.2.2.2-2 c). Unlike regular depletion-mode logic, no level shifting is required, which is an important factor in reducing power dissipation. Actual devices realized in the laboratory with Schottky diode FET logic include NOR gates and ring oscillators. Typical propagation delays are about 140 psec for lower-power devices and 82 psec for high-speed devices (corresponding to equivalent clock rates of about 3.5 and 5.9 GHz, respectively). Power dissipations of 0.335 mW per gate for lower-power devices and 2.5 mW per gate for high-speed devices are typical, yielding speed-power products of 47 fJ and 200 fJ, respectively. Although no figures on processing yields are available, the power consumption of these devices would seem to make them suitable for large-scale integration (LSI; i.e., more than 100 gates/chip).

Low-power MESFETs are made using enhancement-mode FETs. These may be combined to form logic functions just as in the depletion-mode case. Like Schottky diode FET logic, the voltage level at a load is already suitable for driving the next stage (see Figure 1.2.2.2-2 b). Actual devices realized in the laboratory with normally off MESFETs include ring oscillators and basic logic gates. Propogation delays of 290 psec have been reported for these devices (an equivalent clock rate of 1.7 GHz). The lower speed of a normally off MESFET logic is due to the necessity of operating the MESFET in regions where its transconductance ("gain") is small. Power dissipations of 0.17 mW per gate are typical, yielding a speed-power product of 0.05 pJ. The power consumption of these MESFETs would seem to make them suitable for LSI devices with perhaps more than 1,000 gates/chip.

GaAs JFETs (enhancement-mode) work in much the same way as enhancement-mode MESFETs. Some types of logic gates have been fabricated laboratory experiments (see Figure 1.2.2.2-2 d), and work is underway on memory (data) multiplexers and analog-to-digital (A/D) converters. Reported speeds for the logic gates are about 1 nsec. Power consumption is reported to be 2 mW per gate, yielding a speed-power product of 2 pJ. The low power dissipation of GaAs enhancement-mode JFETs gives rise to a promise of LSI logic with nanosecond or subnanosecond propagation delays.

TEDs [sometimes called Transferred-Electron Logic Devices (TELDs)] may also be used for high-speed logic devices. Some devices that have been fabricated in the laboratory using TEDs include a Binary Phase-Shift Keyed (BPSK) modulator, BPSK demodulators, a one-bit A/D converter cell (cascadable for multi-bit conversions), shift registers, and logic gates. Most of these devices actually use FETs as well as TEDs. Reported capabilities of the multiplexers and demultiplexers include top carrier frequencies of 5 to 10 GHz and top data rates of 1 to 1.6 Gbits/sec. Reports for the A/D cell indicate a top sample speed of 5 giga samples/second with a 25-psec aperture. Few data are available for speeds of other devices. Power dissipation of TEDs is very high, with one report of 900 mW per logic gate. Other reports indicate that power consumption figures of 180 to 250 mW per gate are currently achievable and speed-power products of 1 to 2 pJ are achievable in the near future. The power consumption figures alone would limit GaAs TEDs to SSI products at the present time.

Several integrated circuit manufacturers, domestic and foreign, have shown continued interest in GaAs digital devices. This, coupled with the promise of very-low-power and/or very-fast devices and the interest shown by the military, will keep the research in GaAs devices for GHz logic active. It is not evident at this time when production devices will reach the marketplace.

1.2.2.3 Josephson Junctions - Josephson junctions are logic devices built to operate at cryogenic temperatures. Basically, a single junction is two pieces of metal in a superconducting state separated by

an insulator, as illustrated in Figure 1.2.2.3-1. This insulator may be in either of two states, depending on the presence or absence of a magnetic field. This field is controlled by switching a small current through a control line. In the absence of the control current (and therefore, magnetic field) the insulator becomes a superconductor itself and current may flow across it with no voltage drop. In the presence of control current (and therefore, magnetic field) the insulator goes into a state that allows electrons to tunnel through it with a small voltage drop.

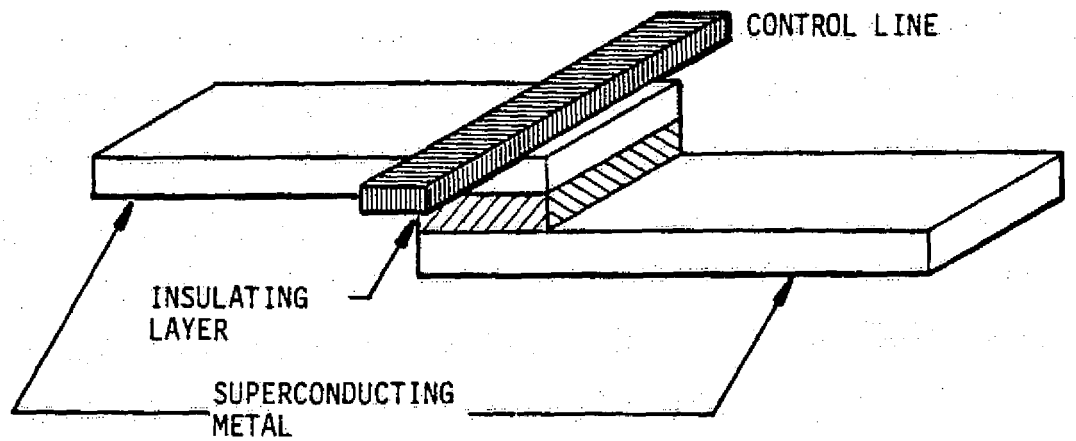


FIGURE 1.2.2.3-1. DIAGRAM OF A JOSEPHSON JUNCTION

Josephson junctions exhibit both high speed and low power dissipation. Furthermore, device sizes are also suitable for LSI to VLSI. However, many problems remain to be overcome before completely practical VLSI Josephson junction devices may be fabricated. For instance, very thin, very uniform insulating layers are required. This problem is close to being solved. Also, mechanical problems resulting from stresses between the several layers of a junction during temperature cycling between cryogenic (near absolute zero) and room temperatures need to be fully solved.

An interesting projection of what may be expected from Josephson junction technology was recently given by Dr. Wilhelm Anacker of IBM (Ref. 1-4). Dr. Anacker's projection, presented in Table 1.2.2.3-1, gives a comparison of the volume, speed, and power of a future Josephson junction computer with an IBM 3033.

TABLE 1.2.2.3-1. COMPARISON OF IBM 3033 PERFORMANCE WITH DR. WILHELM ANACKER'S JOSEPHSON JUNCTION COMPUTER PROJECTION

	VOLUME	SPEED	POWER (LESS COOLING)
IBM 3033*	~4 × 10^5 in^3	6 MIPS	50 kW
IBM Josephson Junction Computer*	8 in^3	300 MIPS	100 mW
Improvement Factor	5 × 10^4	50	5 × 10^5

*Both configurations assume 4 Mbytes of memory.

1.2.2.4 Magnetic Bubbles - Magnetic bubble solid-state memories are discussed in Section 3.2.

1.2.2.5 CCD Devices - CCD solid-state devices are discussed in Subsection 7.2.1.3.

1.2.2.6 Other New Solid-State Technologies - I^2L bipolar solid-state devices are discussed in Subsection 7.2.1.1. MOS solid-state targets used in Electron Beam Addressed Memories are discussed in Subsection 7.2.2.1.

REFERENCES - Section 1

1-1 G. Louie et al., "A Dual Processor Serial Data Controller Chip", 1977 IEEE International Solid-State Circuits Conference, pp. 147-148

1-2 H. Yoshimura et al., "A 64 Kbit MOS RAM", 1978 IEEE International Solid-State Circuits Conference, pp. 148-149

1-3 C. Hammer, "A Forecast of the Future of Computation", AIAA 13th Annual Meeting, Washington, D.C., January 10-13, 1977

1-4 "Josephson Superconductor Has Promise", Computerworld, pp. 39-42, March 6, 1978

BIBLIOGRAPHY - SECTION 1

R. E. Eden et al., "Low Power GaAs Digital ICs Using Schottky Diode-FET Logic", 1978 IEEE International Solid-State Circuits Conference, pp. 68-69

B. M. Welch et al., "Planar GaAs Integrated Circuits Fabricated by Ion Implantation", Technical Digest 1977 International Electron Devices Meeting, pp.205-208

"Static-Induction-Transistor Cell May Up RAM Performance Tenfold", Electronics, p. 64, September 15, 1977

T. Sakai et al., "A 100 PS Bipolar Logic", 1977 IEEE Solid-State Circuits Conference, pp. 196-197

J. Nishizawa et al., "Integrated Logic - Static Induction Transistor Logic", 1977 IEEE Solid-State Circuits Conference, pp. 222-223

T. Sugano, T. Okoski, and M. Watanabe, "Solid State, Synonym for Progress", IEEE Spectrum, pp. 51-56, September 1977

T. H. P. Chang, M. Hatzakis, A. D. Wilson, and A. N. Broers, "Electron-Beam Lithography Draws a Finer Line", Electronics, pp. 89-98, May 12, 1977

R. Sugarman, "New Optical Lithography Equipment Promises Higher Density Memories", Electronic Engineering Times, pp. 1, 13, 15, September 19, 1977

L. Altman and C. L. Cohen, "The Gathering Wave of Japanese Technology", Electronics, pp. 99-122, June 9, 1977

I. E. Sutherland and C. A. Mead, "Microelectronics and Computer Science", Scientific American, pp. 210-228, September 1977

R. Pashley et al., "H-MOS Scales Traditional Devices to Higher Performance Level", Electronics, pp. 94-99, August 18, 1977

R. V. Tuyl and C. Lichti, "Gallium Arsenide Spawns Speed", IEEE Spectrum, pp. 41-47, March 1977

K. Shimotori et al., "Fully Ion Implanted 4096-Bit High Speed DSAMOSRAM", 1977 IEEE Solid-State Circuits Conference, February 16-18, 1977

H. Ishikawa et al., "Normally-Off Type GaAs MESFET for Low Power, High Speed Logic Circuits", 1977 IEEE Solid-State Circuits Conference, February 16-18, 1977

2. SPACE DATA PROCESSING ELEMENTS

Space data processing elements include the hardware and software required to process data and thereby change it to a form more acceptable for data transmission, storage, or use on board, as illustrated in Figure 2-1. The hardware technology elements, including general-purpose computers, distributed processors/multiprocessors, signal processors, and component/packaging technology, are discussed in this section. Hardware technology for onboard data base storage is discussed in Section 3, and spacecraft internal communications elements are covered in Section 4. Onboard software is not covered; however, the software technology discussed in Section 11 has implications to space data processing.

Characteristics that have traditionally distinguished space data processing hardware from other types of computer hardware are reliability, packaging, and power type and consumption. In general, the reliability requirements for space applications impose the need to use a mature technology with established reliability characteristics, a high-reliability parts program, intensive quality controls, and appropriate redundancy.

Packaging of most space hardware differs from avionic hardware in that shock mounting is not utilized and cooling is generally accomplished by radiation and by conduction to the spacecraft structure or coldplate. This implies that, in addition to being able to withstand severe shock and vibration, the internal circuitry of the computer must be packaged to provide good thermal conductivity to the external surfaces for efficient cooling. The Space Shuttle Orbiter is an exception in that most of the avionic hardware, including the IBM AP101A computers, uses forced-air cooling.

Spaceborne electronic equipment normally operates from 28-Vdc power supplied by batteries. The power consumption of the onboard computers is usually limited by the available battery power. Onboard computer processing speed is therefore often restricted to conserve available spacecraft power. Consequently, many spacecraft control functions in the past were implemented by ground command. Currently, the trend is toward onboard, computer-managed, autonomous spacecraft operation. This trend should accelerate as the performance and reliability of spaceborne computers continues to improve.

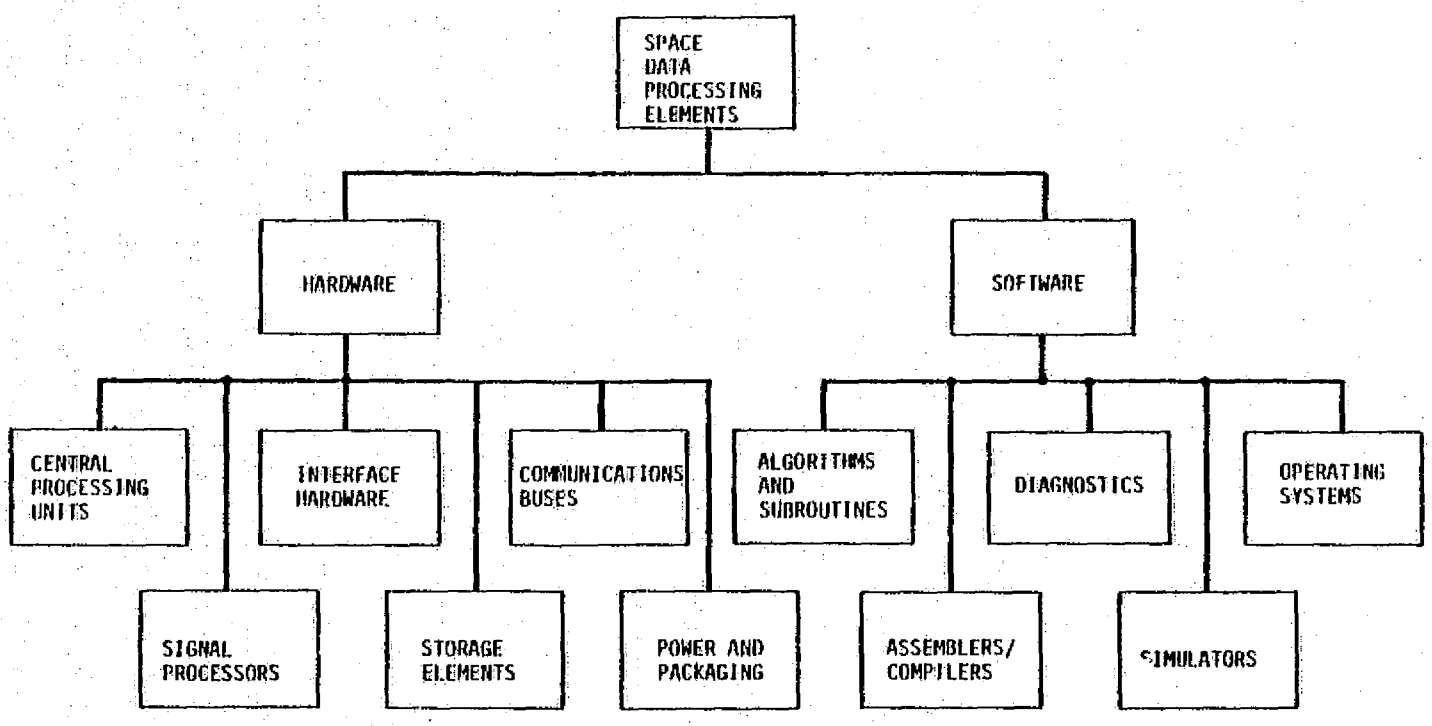


FIGURE 2-1. SPACE DATA PROCESSING ELEMENTS

2.1 GENERAL-PURPOSE SPACEBORNE COMPUTERS

During the past two decades, digital computers have been employed in an increasing variety of spaceborne applications that include guidance, navigation, and control of launch vehicles, manned spacecraft, manned laboratories, and deep space probes; experiment processing on board manned and unmanned spacecraft; and stabilization, pointing, and control of observation and communication satellites. These diverse applications have used a variety of general-purpose computer types ranging in size from small units weighing less than 10 lb with power consumptions under 10 W to large, multiple-processor systems weighing more than 200 lb with power consumptions in excess of 1 kW.

General-purpose spaceborne computers consist basically of various combinations of the following elements:

- CPU
- Memory
- I/O
- Redundancy management hardware
- Power supply
- Mechanical structure or housing.

The discussion in this section places major emphasis on the technology evolution of the CPU and memory elements.

The memory technology is generally the major factor in determining computer size, cost, power, and reliability. Spaceborne computers have traditionally used nonvolatile memory. A nonvolatile memory is one that retains its contents after removal of power. In addition to nonvolatility, some space processor memory designs also provide nondestructive readout (NDRO) operation. In the past, NDRO memory implied the use of plated wire, whereas nonvolatility implied the use of either plated wire or magnetic core for random-access alterable storage. Nonalterable storage, such as used in read-only memories (ROMs) and programmable read-only memories (PROMs), have been used recently in spaceborne computers and represent the first use of semiconductor memories in space.

Techniques for improving the reliability of spaceborne computers include ultraconservative electrical, thermal, and mechanical design and extensive quality controls. To achieve the high reliability necessary for long-term space missions, it is usually necessary to use some form of redundancy to achieve required mission reliability levels. Redundancy methods employed in the past include triple-modular redundancy (TMR) at the component level (Saturn), a standby computer and a watchdog timer (Skylab), two active computers and a spare (IUS), and triple redundant computers. Other redundancy techniques used to improve reliability include the use of redundant power supplies and memories.

Another method for improving the reliability of spaceborne computer memories is the use of bit independent storage coupled with error detecting and correcting techniques. For example, a semiconductor memory can be organized in a "bit-per-BOM" (basic operating memory) configuration with each bit of each word stored in a separate memory. Storing each bit in an independent memory makes it possible for an error-correcting code to be used to correct bit errors. In addition, spare memory modules can be switched in as required to provide even greater improvement in memory system reliability. Such techniques can be used to permit the system to continue operation, even after multiple component failures.

2.1.1 State of the Art in Spaceborne Computers

Spaceborne computer state-of-the-art characteristics include the circuit and packaging technology used to implement the various computer circuit elements, the architecture, and the supporting software. Some of the general characteristics of state-of-the-art spaceborne computer designs are:

- Asynchronous data bus architecture for communication between CPU, I/O, and memory
- Extensive use of highly integrated MSI and LSI logic devices for CPU, memory, and I/O controller implementations
- Modular design for memory and I/O to allow various combinations of standard modules to be easily integrated for different applications
- Communication with other spacecraft subsystems via serial time division multiplexed data buses.

Most of the larger spaceborne computers also provide various hardware and software features or options such as hardware floating point, flexible interrupt handlers, custom I/O controllers, and extensive support software (such as high-level language compilers, simulators, and extensive subroutine libraries). In addition, most offer modular memory designs with various memory module options such as semiconductor, core, and plated wire.

Although there have been some improvements in architectures, the major factor influencing spaceborne computer technology in recent years has been the microelectronics revolution discussed in Sections 1 and 7. Circuit technology improvements have resulted in significant improvements in spaceborne computer speed, power consumption, and circuit density.

At the present time, the architecture of state-of-the-art general-purpose computers can be generally grouped into two categories. The first category includes the smaller microprocessors and microcomputers and the

other category includes all the larger computers. Most of the larger spaceborne computers utilize a CPU that is composed of some type of high-speed bit slice devices (such as the AMD 2900) that are under microprogram control. The microprogrammed bit slice architecture allows the instruction set to be modified by changing the computer microprogram, which is normally implemented as a semiconductor read-only memory.

Several companies have developed space-qualifiable microprocessor/microcomputer LSI hardware based on various semiconductor technologies that include CMOS/SOS, NMOS, and I^2L. Companies that are developing space-qualifiable LSI CMOS/SOS processors include Harris, Hewlett-Packard, RCA, and Rockwell. Hewlett-Packard recently reported the development of a CMOS/SOS 16-bit microprocessor chip. The chip contains 9,600 transistors, executes 34 basic classes of 16-bit instructions with an average instruction rate of 1.05 μsec, and consumes a maximum of 0.5 W when operated from a single 12-V supply.

Integrated Injection Logic (I^2L) microprocessors are currently being built by several companies, including Fairchild and Texas Instruments. Texas Instruments has a ruggedized, 16-bit I^2L microprocessor (SBP 9900) with a complexity of 6,300 gates that operates over a temperature range of -55 to +125°C. The SBP 9900 is software-compatible with the Texas Instruments TMS 9900 microprocessor, which uses N-channel silicon-gate technology. TI has developed a ruggedized, high-reliability family of standard computer modules around the SBP 9900 microprocessor. As with most other microprocessor system hardware, these modules are designed to function either as stand-alone computers or as embedded processors.

Some examples of microprocessors employing NMOS technology include the INTEL 8080, Motorola 6801, Signetics 2650, General Instruments CP1600, Zilog Z-80, Mostek 3870, and the Data General microNOVA; however, only a few of these units, such as the 8080, have been qualified for military or space applications.

Table 2.1.1-1 presents generalized characteristics for state-of-the-art general-purpose spaceborne computers. The characteristics summarized in the table are for general-purpose computers such as used for spacecraft control or onboard data management.

Another way in which microcomputers will be used in future spacecraft data systems is as embedded processors within the various spacecraft subsystem elements to implement specific logical and control functions. As the capabilities of available microcomputers increase, the embedded processors will evolve into general-purpose computers that are contained within the various subsystem elements and function in conjunction with the other spacecraft subsystem processors as a distributed processing system network.

TABLE 2.1.1-1. STATE OF THE ART IN GENERAL-PURPOSE SPACEBORNE COMPUTERS

CHARACTERISTICS	MICROCOMPUTER	LARGE COMPUTER
CPU		
Word Length	8 or 16	8, 16, or 32
Technology	I^2L, NMOS, CMOS/SOS	I^2L
Number of Instructions	40 to 400	70 to 400
Add Time (R-R)	1.5 to 2.0 μsec	1 to 2 μsec
Multiply Time	8 to 12 μsec	3 to 6 μsec
Average Throughput	200 to 300 Kops	500 to 600 Kops
Memory Addressability	64K	1M
Floating Point	Software	Hardware Option
Memory		
Technology	NMOS, Core	NMOS, Core, Plated Wire
Cycle Time		
Internal Capacity	32 Kbytes	128 Kbytes
External Expansion	128 Kbytes	1,024 Kbytes
I/O		
Parallel Rate	200K to 2M bytes/sec	400K to 4M bytes/sec
Serial Rate	1 Mbit/sec	2 to 10 Mbits/sec
System		
Power	20 to 60 W	100 to 350 W
Volume	300 in^3 (32 Kbytes)	1,000 in^3 (128 Kbytes)
Weight	10 to 30 lb	15 to 60 lb
MTBF	30,000 hr	10,000 hr
Type	Binary, 2's Complement Fixed Point	Binary, 2's Complement Fixed or Floating Point

2.1.2 Trends in Spaceborne Computers

The trends in spaceborne computer hardware will closely parallel the trends in microcomputers and minicomputers discussed in Section 7. Spaceborne computer trends will be toward much smaller size, higher processing throughput, and lower power consumption.

Currently, the largest physical items in a large spaceborne computer are the memory and power supply. Current state-of-the-art computer designs generally provide approximately 10^6 bits of storage hardware within the computer housing assembly, in addition to the CPU, I/O, and power supply.

By 1985, space-qualified 10^6 bit random-access memory chips should be available; thus the equivalent storage capacity of a computer such as the IBM AP 101A will be reduced to a single chip. This means that a computer equivalent to the IBM AP101A could be packaged with power supply in a volume of approximately 30 in^3, with most of the space allocated to the power supply and interface circuitry, by 1985. The implications of such a drastic reduction in the physical size of spaceborne computing power goes far beyond the obvious savings in weight, power, cost, etc., and many of the current concepts concerning spacecraft computer system architecture will need to be reexamined.

Because of the significant reduction in CPU and memory size resulting from improved microelectronics, it is anticipated that enough redundancy will be designed into standard spaceborne computer hardware to eliminate all single-point failure mechanisms, thus eliminating many of the problems and the high costs associated with defining and implementing failure-tolerant computers in the past.

The trend in I/O hardware technology will be toward the increased use of optical data base technology to provide extremely high data rates and small size. The optical data bus will also provide extremely good electrical isolation and noise immunity. Optical technology is discussed in Section 4.

2.1.3 Projected Developments in Spaceborne Computers

The advancements in LSI fabrication processes and new device technologies discussed in Section 1 will result in significant improvements in both processor and memory technologies within the next decade. These improvements will produce the following significant changes in spaceborne processors by 1985:

- Faster CPU speeds
- Faster man storage speeds
- Lower power consumption
- Significantly reduced physical size
- Hierarchal memory structure with built-in mass storage.

Table 2.1.3-1 summarizes the projected characteristics for spaceborne computers for 1985.

TABLE 2.1.3-1. PROJECTED TECHNOLOGY FOR SPACEBORNE COMPUTERS, 1985

CHARACTERISTIC	MICROCOMPUTER*	LARGE COMPUTER
CPU		
Word Length	8 to 16	8, 16, and 32
Technology	TBD	TBD
Number of Instructions	100 to 400	200 to 400
Add Time (R-R)	0.1 usec	50 nsec
Multiply Time	0.5 usec	0.3 usec
Average Throughput	10 Mops	20 Mops
Floating Point	Hardware Option	Hardware
Memory (Main)	On CPU Chip	Built-in error correcting
Technology	TBD	TBD
Capacity	32 Kbytes	1 to 4 Mbytes
Mass Storage	Optional	100 Mbytes
I/O		
Parallel Rate	10 Mbytes/sec	20 Mbytes/sec
Serial Rate	10 Mbits/sec	40 Mbits/sec
System		
Power	5 to 10 W	20 to 30 W
Volume*	2 in^3 (less power supply)	30 in^3 (with power supply)
Weight	0.1 lb	2 lb
MTBF	TBD	TBD
Type	Binary, 2's complement Fixed point, redundant CPU	Binary, 2's complement Floating point, redundant CPU

*Will only be used as embedded processors.

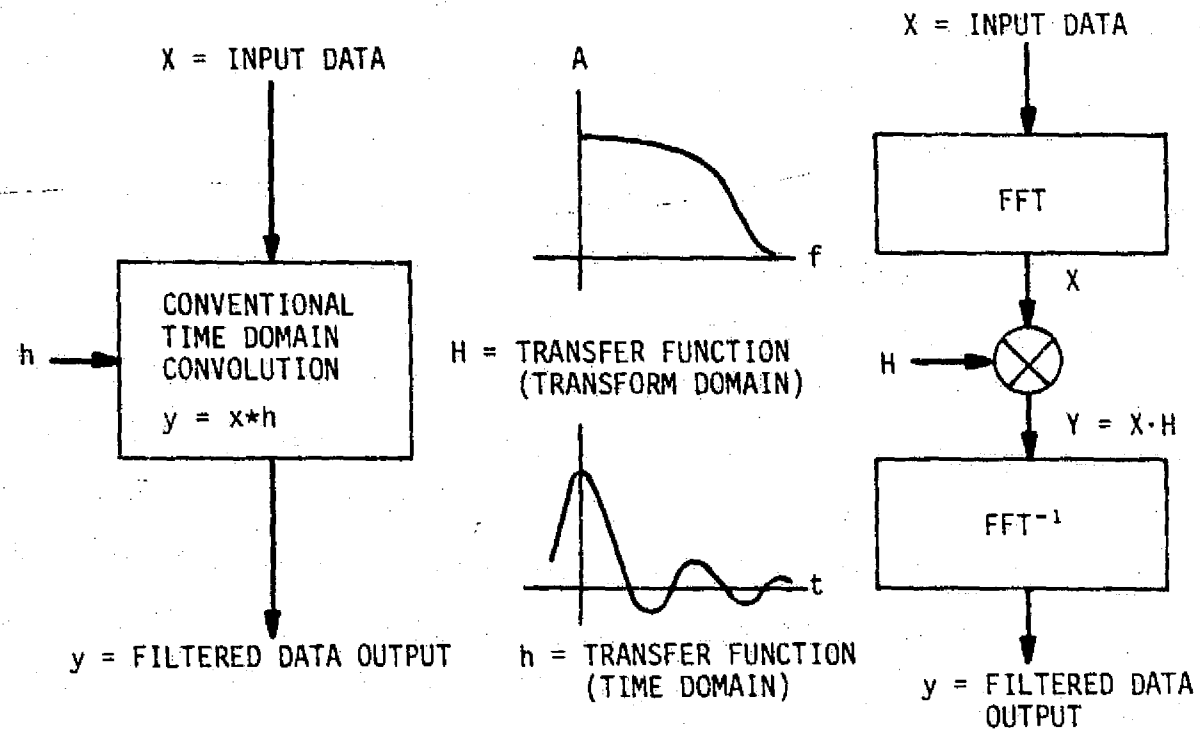
2.2 SIGNAL PROCESSORS

Signal processors are a form of computer that has been architecturally optimized for specific signal processing applications such as digital filtering or performing Fast Fourier Transforms (FFTs). In general, signal processing devices are used to perform a few specific calculations on a repetitive basis and are thus optimized for these specific functions. Some of the more primitive signal processors perform only a few fixed computations and are programmed by hardwired logic. Other types of signal processors utilize either optical or analog computational elements rather than digital logic elements.

Signal processing applications are generally distinguished from general-purpose processor applications by the type of input data and processing required. In most signal processing applications the input data are as a rule very long sequences of data samples such as would be derived from a radar receiver or an optical sensor. Usually these long sequences are segmented into consecutive blocks or arrays for convenience. The distinguishing characteristic of signal processors is not that they operate on arrays of data but rather that they operate on very long sequences of data that may be segmented into arrays.

The FFT and other special transforms are basic to most signal processor applications. For example, consider Figure 2.2-1, which shows two methods for implementing a digital filter. In the time domain, the filter algorithm can be performed by convolving the transfer function of the filter with the input data. In addition to performing convolution in the time domain, the filter can be implemented in the frequency domain by transforming the input function into the frequency domain, performing the filter function, and then transforming the resulting function back into the time domain. This implementation is based on the fact that convolution in the time domain ($y = x*h$) is equivalent to multiplication in the frequency (or transform) domain ($Y = X \cdot H$).

For digital signal processor applications, the above filter function can be implemented in the transform or frequency domain with



a. Time Domain Filter

b. Transform Domain Filter

FIGURE 2.2-1. TWO METHODS FOR IMPLEMENTING DATA FILTER

significantly fewer processor instructions than in the time domain. For example, the time correlation of two complex arrays each containing 1,024 sample points requires more than a million complex multiplies using direct convolution in the time domain and only 16,384 complex multiplies using the FFT approach.

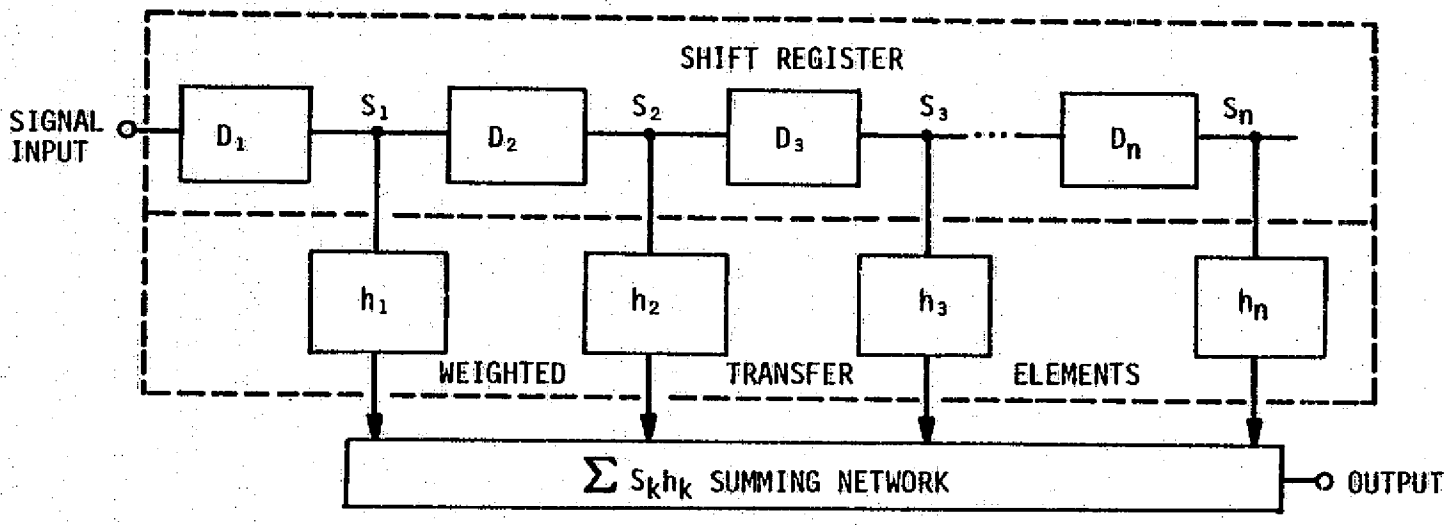
2.2.1 State of the Art in Spaceborne Signal Processors

Signal processor technology is currently proceeding in a number of directions corresponding to various applications requirements, processing techniques, and component technologies. Some recent developments include:

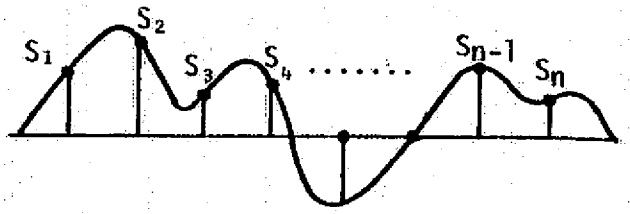
- Analog transversal filters employing CCD technology
- Digital charge-coupled logic, a new digital logic that is particularly well suited for signal processor applications
- The programmable digital signal processor, which represents a new kind of computer that is more specialized than a general-purpose computer yet provides extremely high throughput for certain classes of problems
- The integer ring transform, which is computationally more efficient than the FFT for certain classes of applications.

Analog CCD transversal filters are beginning to find widespread application. For example, Texas Instruments has developed CCD transversal filters for the Synthetic Aperture Radar (SAR) processor that JPL is currently developing, and Reticon Corporation recently announced the first commercially available unit, designated the R5602. The Reticon device is a 64-tap split-electrode device that can be factory tailored to provide various analog filter characteristics.

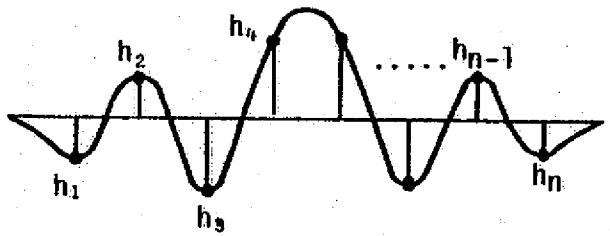
Figure 2.2.1-1 functionally illustrates the basic transversal filter organization consisting of a sampled data shift register, weighted transfer elements, and a summing network. Since the shift register is basically an analog sampled-data delay line, CCD technology is well suited for implementing the shift register portion of the filter. After each clock pulse, the output of each stage is nondestructively multiplied by a weighting factor, h_k. These weighted outputs are then summed to provide the cross-correlation function or the filter output.



TRANSVERSAL FILTER BLOCK DIAGRAM (ANALOG OR DIGITAL)



TYPICAL SIGNAL WAVEFORM



TYPICAL TRANSFER FUNCTION

FIGURE 2.2.1-1. TRANSVERSAL FILTER

The method for generating the fixed weighting function can be understood by referring to Figure 2.2.1-2, which shows a top view of the electrodes in the CCD shift register. The weighting is accomplished by dividing the electrodes of one of the three phases normally used to clock the shift register into two parts as shown. Since the drive current passing through each electrode is directly proportional to the charge being transferred, a gap located in the center will produce no net differential current, or a weighting function equal to zero. As the gap is moved up or down, the corresponding weighting function becomes positive or negative. By connecting all the split electrodes to a common driver as shown and measuring the combined differential current, the transversal filter function of Figure 2.2.1-1 is achieved.

Although CCD analog filters have performance limitations when compared with digital filters, some of their potential advantages include lower cost, smaller size, less power, and higher reliability than digital filters.

Another recently developed device technology that is particularly well suited for spaceborne signal processor applications is Digital Charge-Coupled Logic (DCCL). TRW Systems Group has recently reported the implementation of digital logic functions using charge-coupled devices in pipeline configurations (Refs. 2-1 and 2-2). DCCL provides higher functional density and lower power dissipation than either NMOS, CMOS, or I^2L over the full range of clock frequencies.

Arithmetic functions implemented in DCCL must be organized in a pipeline manner as illustrated in Figure 2.2.1-3 since DCCL does not have the ripple-through logic capability that is needed for carry propagation in binary adders employing conventional logic technologies. Although the DCCL pipeline approach is not suited for general-purpose processors that execute random instructions, it is well suited for signal processing applications such as FFT implementations. An advantage of the DCCL pipeline adder is a very high throughput rate. For example, data enter the adder at the maximum clock rate (two input words each cycle), and sums exit the adder some time later at the same maximum clock rate.

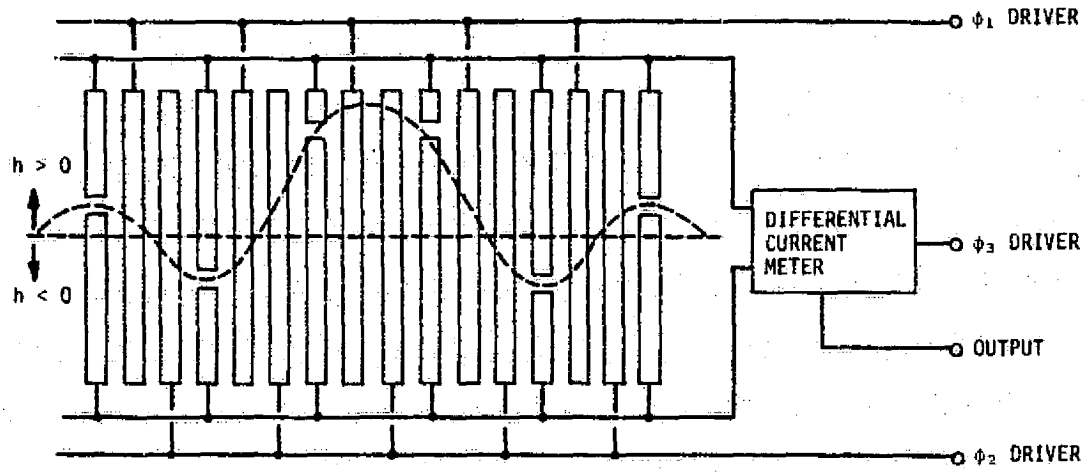


FIGURE 2.2.1-2. CCD ELECTRODE WEIGHTING ILLUSTRATED FOR ANALOG TRANSVERSAL FILTER

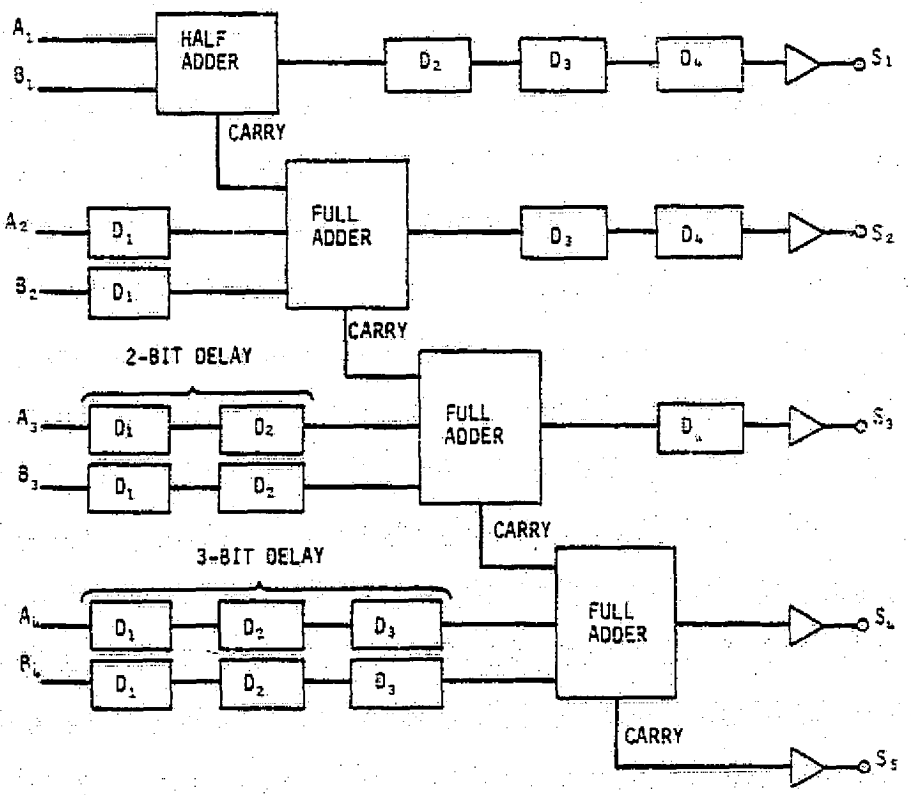


FIGURE 2.2.1-3. BLOCK DIAGRAM OF DCCL 4-BIT PIPELINE ADDER

It should be noted that the eventual impact of DCCL will not be in the production of single functions such as a 16 by 16 multiplier. The major impact will be in the development of complex VLSI signal processing devices that contain multiple multiplier/adder subfunctions.

The state-of-the art programmable digital signal processor is a form of general-purpose processor that has been optimized for signal processing applications. The present machine has evolved as a result of efforts to combine the efficiency of a hardwired signal processor with the computational flexibility of a general-purpose computer. Some typical state-of-the-art architectures are described in References 2-3, 2-4, and 2-5. In general, all of these configurations employ some form of functional parallelism; i.e., they use a parallel array of lower performance devices (arithmetic units and memory units) to achieve a higher performance function. Another characteristic of state-of-the-art programmable digital signal processors is the separation of the overall control, timing, and test functions from the memory and computational elements. This division allows a single host or control processor to provide parallel control of several array processors.

IBM in Reference 2-6 has briefly described the implementation of an integer ring transform that is reported to be much faster than the FFT for the class of applications to which it applies. According to IBM, the integer ring transform requires less than 1,024 complex multiplies to perform a cross-correlation of two 1,024-point arrays that normally requires 16,384 complex multiplies using the FFT approach.

2.2.2 Trends in Spaceborne Signal Processors

The development of signal processing technology has been supported primarily by the real-time processing requirements of military radar, sonar, and electronic countermeasure (ECM) systems. Technology is currently advancing in the areas of transform algorithms, device technology, architecture, and software.

Architecturally, signal processors are being designed to provide greater signal processing throughput and flexibility at reasonable costs. This is being accomplished by combining hardware and software modularity with top-down programming in a way that achieves very high signal processing throughput while allowing the system to be programmed efficiently. Some of the current trends in signal processor hardware and software are the following:

- Evolution of faster, more complex, and lower power LSI and VLSI signal processing chips
- Architecturally, the trend is toward a central control function with multiple arithmetic and memory units. The arithmetic units are typically structured in a "pipeline" organization.
- Software trends are generally following existing trends in the general-purpose computer area; i.e., high-order languages are evolving to achieve greater programmer efficiency.
- The development of new computational techniques (transform algorithms) that increase processing speed without increasing computer operational requirements.

The combined effect of these trends should result in significantly greater throughput for low-power programmable spaceborne signal processors by 1985, as illustrated in Figure 2.2.2-1.

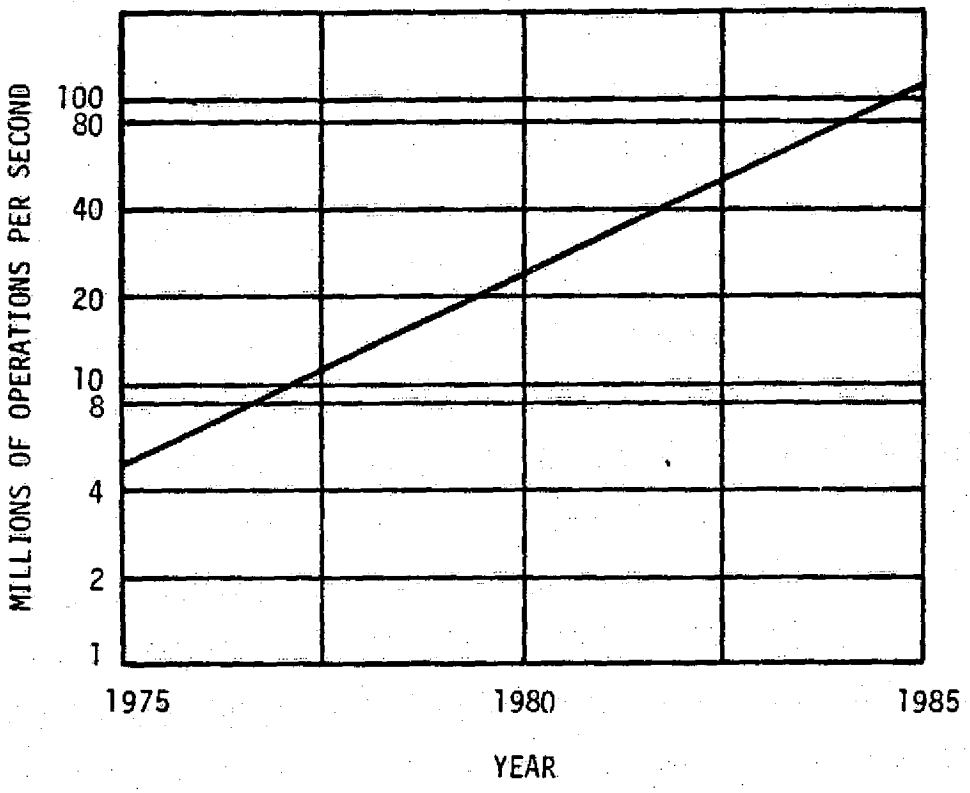


FIGURE 2.2.2-1. TRENDS IN PROGRAMMABLE SPACEBORNE SIGNAL PROCESSOR THROUGHPUT

2.2.3 Projected Developments in Spaceborne Signal Processors

The development of new signal processing computational techniques based on recent theoretical developments is expected to have a marked impact on future signal processor capabilities. The new approaches rely heavily upon number theory and advanced algebraic concepts such as ring theory and Galois field theory. These techniques generally involve converting the input data into a residue arithmetic representation, performing all calculations with respect to small integer moduli, and finally, using the Chinese Remainder Theorem (CRT) to decode the results. Some of the promising new transform techniques include the Number Theoretic Transform and the Galois Field Transform. The advantages of these new transform techniques coupled with faster logic technologies and faster architectural implementations should press airborne/spaceborne signal processing speeds to around 100 million operations per second by 1985.

Signal processor architectures are generally distinguished by the way in which they implement multiple simultaneous computations as opposed to general-purpose computers that typically execute only one operation at a time. At the present time, there exists very little commonality in programmable signal processor architectures; however, it is anticipated that by 1985, families of software-compatible signal processors will have evolved, supported by high-level languages for efficient development programming. Another factor that will influence the trend toward greater hardware and software commonality is the increasing application of signal processors in commercial applications, such as seismic analysis and mechanical vibration analysis.

REFERENCES - SECTION 2

2-1 T. A. Zimmerman, et al., "Digital Charge-Coupled Logic (DCCL)", IEEE Journal of Solid-State Circuits, Vol. SC-12, No. 5, pp. 473-484, October 1977

2-2 "New CCDs: the Least Power and the Highest Density", Electronic Design, p. 19, October 25, 1977

2-3 W. R. Loutensock, et al., "A Distributed Programmable Signal Processor and Forecast of Performance in the Eighties", The MSFC/UAH Data Management Symposium, pp. 283-288, October 18-19, 1977

2-4 S. F. Hsueh, et al., "Future Aerospace Digital Signal Processing Concepts", AIAA/NASA/IEEE/ACM Computers in Aerospace Conference, pp. 75-81, Los Angeles, California, October 31 - November 2, 1977

2-5 W. E. Arens, "CCD Architecture for Spacecraft SAR IMAGE Processing", AIAA/NASA/IEEE/ACM Computers in Aerospace Conference, pp. 75-81, Los Angeles, California, October 31 - November 2, 1977

2-6 R. L. Carberry, "Trends in Aerospace Computers", AIAA/NASA/IEEE/ACM Computers in Aerospace Conference, pp. 7-10, Los Angeles, California, October 21 - November 2, 1977

BIBLIOGRAPHY - SECTION 2

L. G. Dang, et al., "A CMOS/SOS 16-Bit Parallel μ CPU", ISSCC Digest of Technical Papers, pp. 134-135, 245, February 1977

M. Fernandez, "A Review of DOD and NASA Computer Standardization", AIAA/NASA/IEEE/ACM Computers in Aerospace Conference, pp. 232-239, Los Angeles, California, October 31 - November 2, 1977

"Self Healing Computer in Development", Aviation Week and Space Technology, pp. 57-60, August 15, 1977

S. E. Ozga, "CMOS/SOS Processors", AIAA/NASA/IEEE/ACM Computers in Aerospace Conference, Los Angeles, California, October 31 - November 2, 1977

B. Connolly, "Microprocessors to get mil specs", Electronics, p. 76, September 1, 1977

"Microcircuit Speed Increase Sought", Aviation Week and Space Technology, pp. 59-60, October 24, 1977

D. Huffman, et al., "Substrate Bias Generator Optimizes Military MOS Memories", Military Electronics/Countermeasures, pp. 62, 64, 76, September 1977

C. Erickson, et al., "A 16-Bit Monolithic I^3L Processor", 1977 ISSCC Digest of Technical Papers, pp. 140-141, 247, February 17, 1977

3. SPACE DATA STORAGE ELEMENTS

Space data storage elements discussed in this section pertain to the devices and techniques used to store space data for either delayed transmission, delayed processing and display onboard, or physical return of the storage media. Figure 3-1 identifies the subelements that were considered as potential candidates for such storage. The majority of the subelements identified in the referenced figure have been ruled out as likely candidates for space data storage either because certain features of these subelements are undesirable for a space data storage system or because they have not been used previously, and there are no known developments going on to make them more attractive for space applications.

The most likely candidates for space data storage between 1977 and 1985 appear to be tape recorders (currently the primary means of data storage for subsequent playback to Earth), magnetic bubbles, and film (for return to Earth physically). Tape recorders are addressed in Subsection 3.1, and bubble memories are addressed in Subsection 3.2.

The approaches that have been ruled out as being likely candidates for space data storage and the rationale for their elimination are:

- Disks - Disks had been ruled out previously in favor of tape recorders, and the technology improvements occurring in both areas do not tend to give disks any more favorable consideration than in the past.

- Semiconductors - Semiconductors have been ruled out primarily on the basis of the volatility. However, certain semiconductor devices (CCDs for example) have a number of desirable features that could make them contenders for future use as space data storage devices. Also, certain types of semiconductors (MNOS) do not exhibit the characteristic of volatility. Semiconductor memory technology is discussed in Section 7.2.1.

- Core - Core was eliminated from contention on the basis of cost, speed, and size.

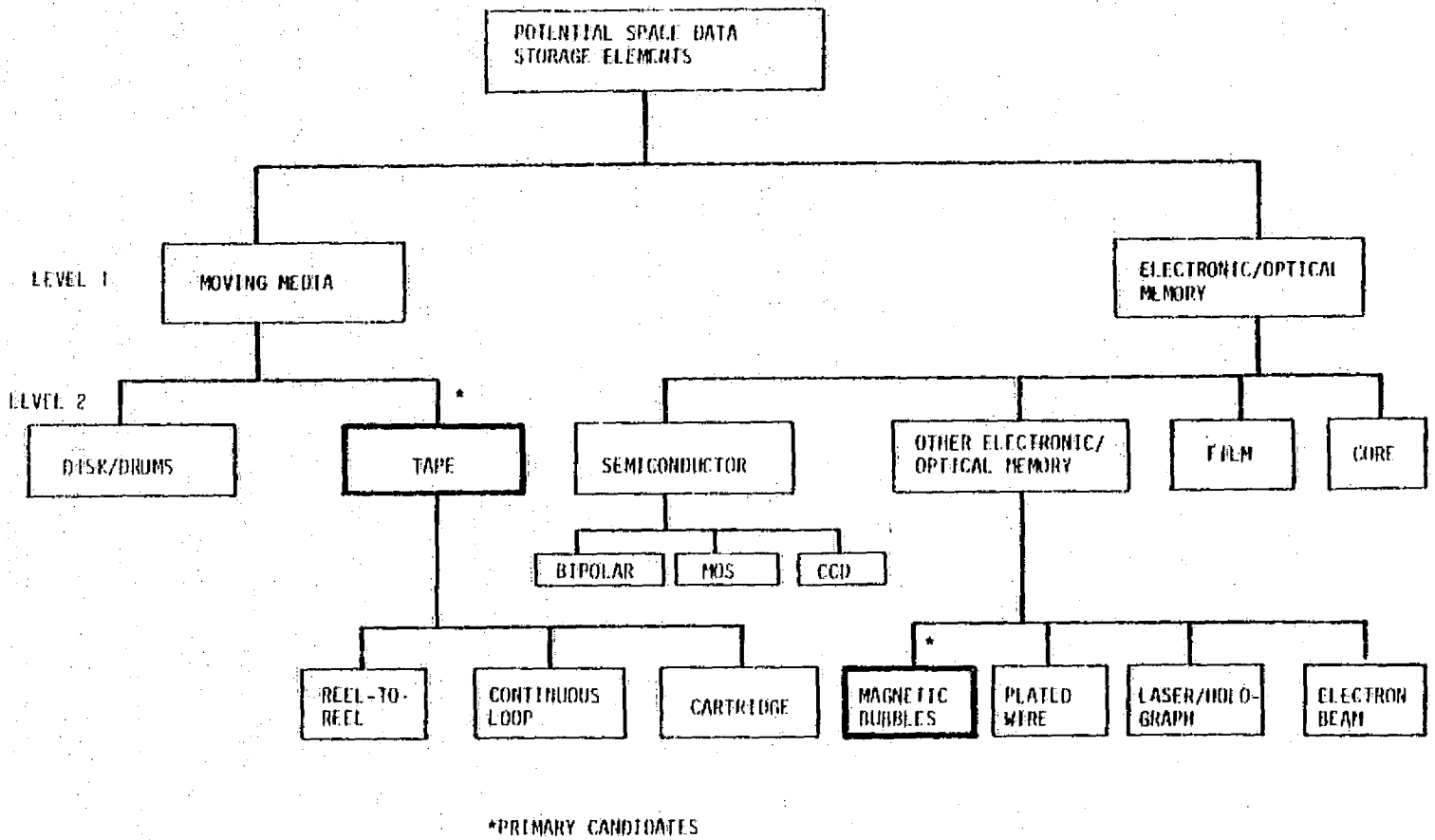


FIGURE 3-1. POTENTIAL SPACE DATA STORAGE ELEMENTS

- Plated Wire - Plated wire was not considered a likely candidate on the basis of storage capacity and cost.

- Optical Systems - Optical systems were eliminated on the basis of technology status, which is primarily limited to laboratory type systems at present.

- Electron Beam Addressable Memories - E-beam memories were ruled out because ongoing development work on this technology is limited, even for ground applications. Other technologies appear capable of achieving the same effect as E-beams with less development costs.

3.1 AEROSPACE MAGNETIC TAPE RECORDERS

Magnetic tape recorders have provided the primary means for mass data storage onboard spacecraft since the early days of the space program, and at the time of this report, they are still the only operational system capable of being used for storing and retrieving dynamic data in space. This section is restricted to tape recorders that are qualified for aerospace applications. Section 6 addresses wideband tape recorders of the type used for recording wideband instrumentation/telemetry data; Section 7 addresses digital tape recorders for data processing applications; and Section 8 addresses mass storage tape recording devices of the type used for archival storage. High-performance magnetic tape recorders are generally classified as either digital recorders, instrumentation recorders, or video recorders. Digital tape recorders are used primarily to support data processing applications and are characterized by the need for start-stop operation to permit real-time access and readout of individual records within a tape file. Instrumentation recorders generally are capable of higher performance in terms of storage density and throughput than data processing tape recorders. They are also characterized by continuous rather than incremental operation. Instrumentation recorders are used primarily to store wideband data, including wideband analog and digital telemetry signals, both on the ground and onboard spacecraft. Video recorders are extremely wideband recorders designed for recording and replaying video signals such as those associated with broadcast television. The video recorders achieve extremely wide bandwidth by utilizing rotating heads to effectively increase the speed at which the tape moves across the head. Video recording techniques are currently being applied to the design of both instrumentation and digital tape recorders to achieve higher storage capacities and higher data transfer rates.

The emphasis throughout this section is on recording densities, data rates, and capacities. These parameters are a function of the record head and the recording media characteristics. Although not discussed in this section, significant improvements in the state of the art are also occurring in the recorder electronics. These improvements are discussed in other sections of the report.

3.1.1 State of the Art in Aerospace Magnetic Tape Recorders

Aerospace magnetic tape recorders have traditionally used either direct, frequency modulation (FM), or high-density digital recording techniques for recording onboard data. The recorders that provide the highest performance and represent the state of the art in technology use digital high-density recording techniques. Thus this portion of the report is directed toward that class of recorder.

The major improvements in aerospace tape recorder designs in recent years have centered around increased data transfer rates and increased storage capacity. In both cases, the performance improvements have been achieved primarily by increasing the number of tracks per inch, either by using higher-density longitudinal heads or by employing rotating-head video recording techniques.

The ERTS (Landsat), Skylab Airlock, and Apollo/Soyuz recorders all used rotary-head recorders. Recording densities of 6.25×10^5 bits/in^2 were achieved on these programs using this technique. The NASA 240-Mbit/sec recorder currently under development employs longitudinal (fixed-head) recording techniques and will be capable of achieving a storage density of 2.7×10^6 bits/in^2. The U.S. Navy is currently developing a variable-data-rate, rotary-head recorder with a storage density of 2.7×10^6 bits/in^2. Although the Navy unit is not an aerospace recorder, it illustrates what can currently be achieved with ruggedized hardware in terms of storage density. Specifications for some of the various NASA aerospace tape recorders currently under development are presented in Table 3.1.1-1.

Technology advances are taking place in a number of areas of magnetic tape recording that will affect the future capabilities of spaceborne high-density digital recorders. Among the most important advances are improvements in:

- Read-write head materials
- Read-write head fabrication techniques
- Automatic scan tracking techniques
- Continuously variable data rate recording techniques
- Tape materials
- Channel coding techniques.

TABLE 3.1.1-1. NASA AEROSPACE TAPE RECORDERS CURRENTLY UNDER DEVELOPMENT

MANUFACTURER	ODETICS	ODETICS	RCA
Model/Name	SCS 1001 (NASA 10^9)	DDS 6000 (Spacelab HDRR)	240 MBS
Year Operational	1978	1978	1979-80
Volume	1,648 in^3	3,393 in^3	5.3 ft^3
Weight	TBD	85 lb	200 lb
Maximum Input Rate	8.192 Mbits/sec	32 Mbits/sec	240 Mbits/sec
Maximum Record Speed	120 in/sec	92 in/sec	100 in/sec
Number of Tracks	8 or 16	22 data + 2 servo	120 data + 22 other
Tape Length	2,442 ft	9,200 ft	6,200 ft
Tape Width	0.5 in.	1.0 in.	2 in.
Storage Capacity	4×10^9 (16 Tracks)	3.8×10^{10}	1.8×10^{11}
BER	5×10^{-7}	1×10^{-6}	1×10^{-6}
Head Type	Fixed	Fixed	Fixed
Configuration	Coaxial	Coaxial	Coplanar
Maximum Power	31W (serial)	161W	270W

Several companies have reported improvements in head materials, fabrication, and design techniques. These include the multi-turn, thin-film head design recently reported by Applied Magnetics, Inc., and improvements in rotary-head designs, such as IBM's 3850 helical-scan digital recorder in which the heads "fly" at a nominal 10 μin. above the tape surface. Because rotating-head recorders have tape-head speeds of 1,000 to 1,500 in/sec, a head life of about 1,000 hr is considered normal with "in-contact" heads. The "flying" head should provide virtually unlimited head life.

Improvements in automatic scan tracking allow the reproduce head in video tape recorders (VTRs) to follow the written track more accurately and thus allow closer track spacings. State-of-the-art VTRs achieve tracking errors of less than 1 mil through servo control of the tape speed and drum speed and through mechanical guidance of the tape. Similar techniques are being developed for high-density, fixed-head recorders. Automatic scan tracking also makes possible continuously variable data rate changes.

Continuously variable data rate recording provides a high degree of adaptability for data systems with variable data rates. These types of recorders should find increased application in future spacecraft with multiple observation systems that operate asynchronously.

The development of dc-free modulation codes for recording is an important improvement in digital instrumentation recorder design because it provides a significant improvement in BER without an increase in bandwidth. Typically, repetition of a worst-case pattern with large dc content in a conventional digital recorder causes a dc buildup that results in a distorted output waveform and increased BER. The new codes are adaptive such that a code changes patterns to eliminate dc buildup whenever a pattern with the potential of dc buildup occurs.

3.1.2 Trends in Aerospace Magnetic Tape Recorders

The trends in aerospace magnetic tape recorders are toward more tracks per inch of tape width, use of wider (2-in.) tapes, higher data transfer rates, faster start/stop times, variable data rate recording higher performance tapes, and use of more sophisticated coding for error control. Other trends include greater use of integrated circuitry, and increasing use of rotary-head techniques. A resulting trend is lower cost per bit of storage.

The trend toward more tracks per inch of tapewidth is driven by efforts to achieve greater packing densities (i.e., higher bits per square inch and higher bits per cubic inch) and higher data rates. The reason for increasing tracks per inch (tpi) rather than bits per inch (bpi) is that for a moderately high-bpi system, a bpi-versus-tpi tradeoff favors tpi increase and bpi decrease. The reason for this is that a 2:1 increase in bpi packing density costs 18 dB of S/N margin, whereas increasing the tpi by 2:1 costs only 3 dB. Difficulties with multi-channel head fabrication and with multi-channel electronics have so far prevented widespread adoption of fixed-head machines with track densities exceeding 42 tpi.

Rotary-head digital recording is emerging as probably the most viable magnetic tape recording technique for future spacecraft storage applications. Rotary-head recorders have already been used for at least three different spacecraft data recording applications and are also finding increased application as ground-based instrumentation recorders. This trend, plus the need to provide both digital recording with fast start-stop operation and onboard mass storage capability onboard future spacecraft, assures the increased use of rotary-head recorders in space. One of the main advantages of the rotary-head recorder is its ability to provide fast start/stop operation with minimum tape motion.

Currently, high-data-rate (>100 Mbits/sec) recording can only be accomplished by using longitudinal fixed-head type recorders since the maximum data rate is a linear function of tape speed, number of tracks, and bits per inch of tape. Existing rotary-head type recorders do

not employ multi-track heads and thus are limited to approximately 20 Mbits/sec. Rotary-head recorders currently under development will employ multitrack heads and provide I/O rates in excess of 100 Mbits/sec.

Authorities in the field predict that both rotary-head and fixed-head machines with transfer rates of 600 Mbits/sec using 2-in. tapes will be available by the early 1980's. Figure 3.1.2-1 illustrates the trend in maximum data transfer rates for spaceborne tape recorders through 1985. As may be seen in this figure, the data transfer rate is projected to increase to 1 Gbit/sec by 1985. The technology currently exists to build 1-Gbit/sec ground-based recorders; however, the limited market does not justify the development cost. By 1985, the widespread use of 1-Gbit/sec optical data links should create a significant market for high-rate recorders for ground-based applications.

Figure 3.1.2-2 illustrates how the number of tracks per inch is expected to increase through 1985 for both fixed-head and rotary-head recorders. As previously explained, the increase in track density will be made possible by a combination of head fabrication techniques and automatic tape-track-following techniques using a servo loop.

Figure 3.1.2-3 shows the future trend in areal storage density for both aerospace recorders and ground recorders. Although there are no technical reasons, other than power and weight requirements, for the aerospace recorders to lag the ground-type recorders in areal storage density, most of the current developments in state-of-the-art recorders are for either ground-based or shipboard applications. For example, RCA is developing a 2.7-Mbit/in^2 recorder for the Navy, and Bell and Howell has a contract with Northrop to develop a 270-Mbit/sec recorder for the Air Force, both of which exceed NASA airborne recorder capabilities. One source indicated that NSA is currently looking for a 600-Mbit/sec recorder; data on this recorder were unavailable.

Another current trend is to develop and space-qualify bubble memory systems as replacements for spaceborne tape recorders. Actually,

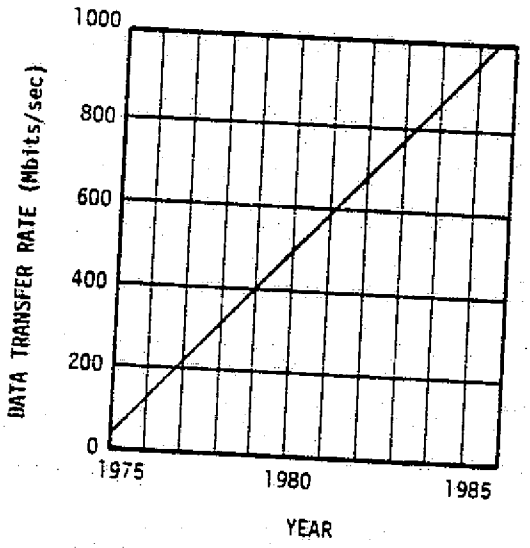


FIGURE 3.1.2-1. PROJECTED MAGNETIC TAPE DATA TRANSFER RATES AS A FUNCTION OF TIME FOR 2-in.-WIDE TAPES

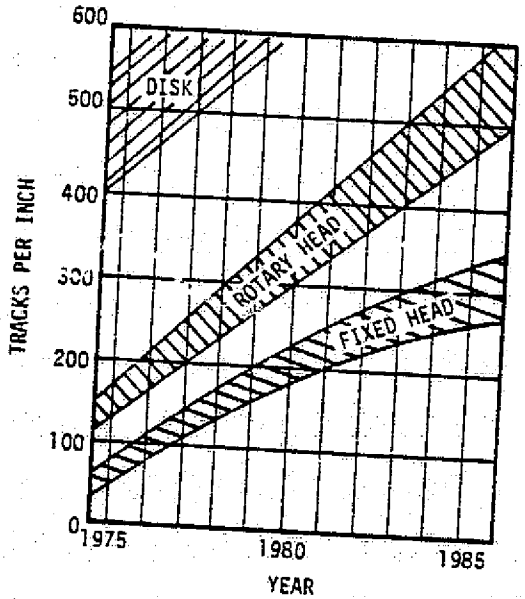


FIGURE 3.1.2-2. MAGNETIC TAPE LINEAR DENSITY AS A FUNCTION OF TIME FOR FIXED-HEAD AND ROTARY-HEAD RECORDERS

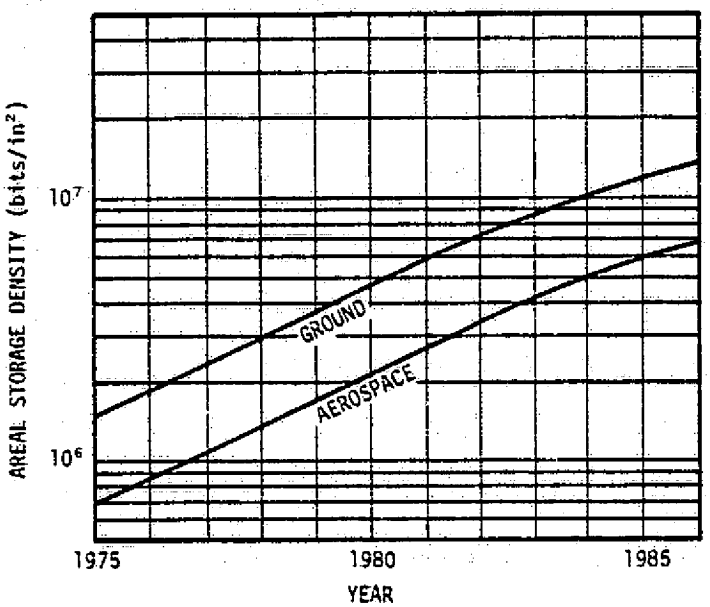


FIGURE 3.1.2-3. PROJECTED MAGNETIC TAPE STORAGE DENSITY AS A FUNCTION OF TIME

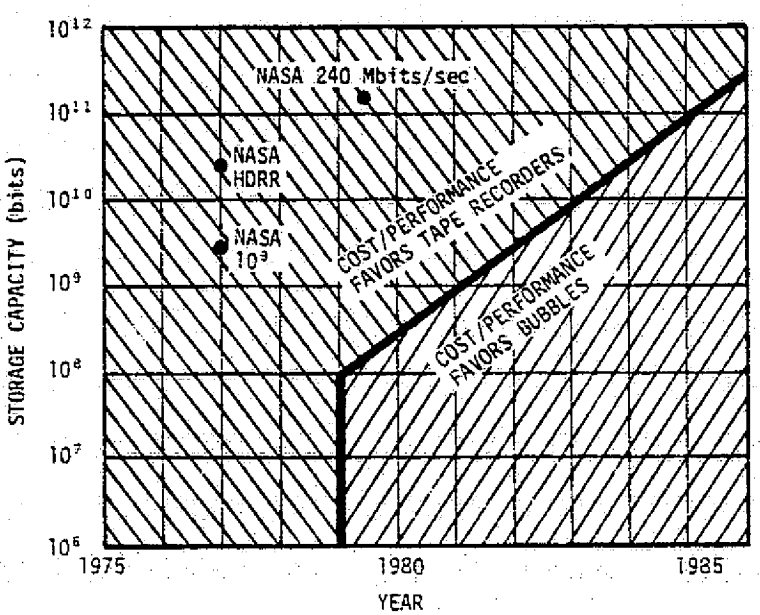


FIGURE 3.1.2-4. COST/PERFORMANCE CROSSOVER POINT AS A FUNCTION OF TIME FOR MAGNETIC TAPE RECORDERS AND MAGNETIC BUBBLES

these units will only be used to replace relatively low-capacity recorders, but eventually they will be competitive for the storage applications up to 10^{12} bits. Figure 3.1.2-4 shows the trend in the cost/performance crossover point for magnetic tape versus bubbles through 1985. The 10^{8} bit recorder presently under development was assumed as the starting point for this curve. The crossover point as a function of time is based on the use of approximately 1,000 state-of-the-art magnetic bubble chips (see subsection 3.2) to build a recorder. The cost for a magnetic bubble recorder will still be more than a comparable capacity tape recorder, but the reliability and lifetime will be much greater. Based on the curve presented in Figure 3.1.2-4, NASA will have at least a 5-year lifetime after development for the tape recorders presently being developed before being replaced by an equivalent bubble memory system.

3.1.3 Projected Developments in Aerospace Magnetic Tape Recorders

Advancements in aerospace tape recorders will closely parallel the advancements in ground-based tape recorders, as previously discussed. Future spaceborne tape recorder developments will be limited to those providing extremely high data storage capacity. The requirements for recorders of low data capacity will be filled by magnetic bubble memory systems as shown in Figure 3.1.2-4.

Aerospace tape recorder technology in the 1985 timeframe should provide storage densities approaching 6×10^6 bits/in^2 and provide data transfer rates of up to 10^9 bits/sec. The projected characteristics of a high-performance spaceborne recorder in the 1985 timeframe are given in Table 3.1.3-1. In addition to very high-capacity data recorders, it is anticipated that the increased application of spacecraft onboard processing will result in the need to develop a large-capacity spaceborne mass storage system that is analogous to the mass storage systems discussed in Section 8. These mass storage systems will probably be implemented in the form of automatic tape libraries with intermediate bubble storage for staging and data buffering. Technology will support the developments listed, but the requirement for spaceborne mass storage systems will depend heavily on the emphasis of NASA programs in the mid-1980's.

TABLE 3.1.3-1. PROJECTED CHARACTERISTICS FOR AEROSPACE MAGNETIC TAPE RECORDERS IN 1985

Storage Capacity	10^{12} bits
Tape Width	2 in.
Tape Length	8,400 ft
Maximum Data Rate	1 Gbit/sec
Maximum Recording Speed	100 in/sec
Weight	200 lb
Size	5 ft^3
BER	1×10^{-6}
Maximum Power	200 W

3.2 MAGNETIC BUBBLE MEMORIES

Memory devices that employ magnetic bubble technology exhibit very favorable characteristics for space data storage applications, including high storage densities, large data storage capacities, relatively low power consumption per bit of storage, high reliability, and nonvolatility. In addition to space data storage applications, magnetic bubble technology is considered the front-runner in the race to fill the speed/capacity/cost gap that exists between semiconductor memory systems and rotating magnetic mass memories.

Magnetic bubbles are mobile cylindrical magnetic domains that can be generated in a thin, flat magnetic film material and moved about in the medium by the application of an external magnetic field. The bubbles are guided within the film layer (storage array) by interaction with propagate paths that are processed into the layer and magnetized by the external magnetic field. The process by which bubbles move along the propagate path in response to a magnetic field is illustrated in Figure 3.2-1. A bit of information is represented in this figure by the presence or absence of a magnetic bubble. (The presence of a bubble may be interpreted as a "1", in which case the absence of a bubble would represent a "0".) During the write process, the chip circuitry generates the bubble and handles the transfer into the storage area. During the read process, the circuitry performs a replicator/annihilator function whereby bubbles are duplicated, with the copy going to the detector and the original being returned to the storage area. Chip circuitry also provides the detection function, whereby either the presence or absence of a bubble is sensed.

Magnetic bubble memory technology has advanced dramatically since 1967, when magnetic bubbles were first invented by A. H. Bobeck at Bell Laboratories. The first devices employed 100-μm bubbles and provided a storage density of 680 bits/cm^2. By contrast, current laboratory devices employ 1.8-μm bubbles and provide storage densities of 1.6×10^6 bits/cm^2. Impressive as this past work has been, future work may be even more impressive as bubble diameters tend toward the theoretical limit, which is estimated to be 0.01 μm (Ref. 3-1). Geometric

features and the effects of these features on performance characteristics of magnetic bubble storage devices are discussed in subsequent paragraphs.

Magnetic bubble storage systems capable of providing large-capacity storage for space applications are being developed by both NASA and the U.S. Air Force. The Air Force efforts include a program with Texas Instruments to develop a fast-access mass memory, support of second-generation technology employing smaller bubbles, and a feasibility study with Rockwell for a 10^{12}-bit airborne memory. NASA is currently developing space-qualifiable 10^{7}- and 10^{8}-bit bubble memory systems for space data storage applications. This development work, combined with the rapid developments in the commercial arena, makes magnetic bubble memory a primary contender for future use as a mass storage device in spacecraft.

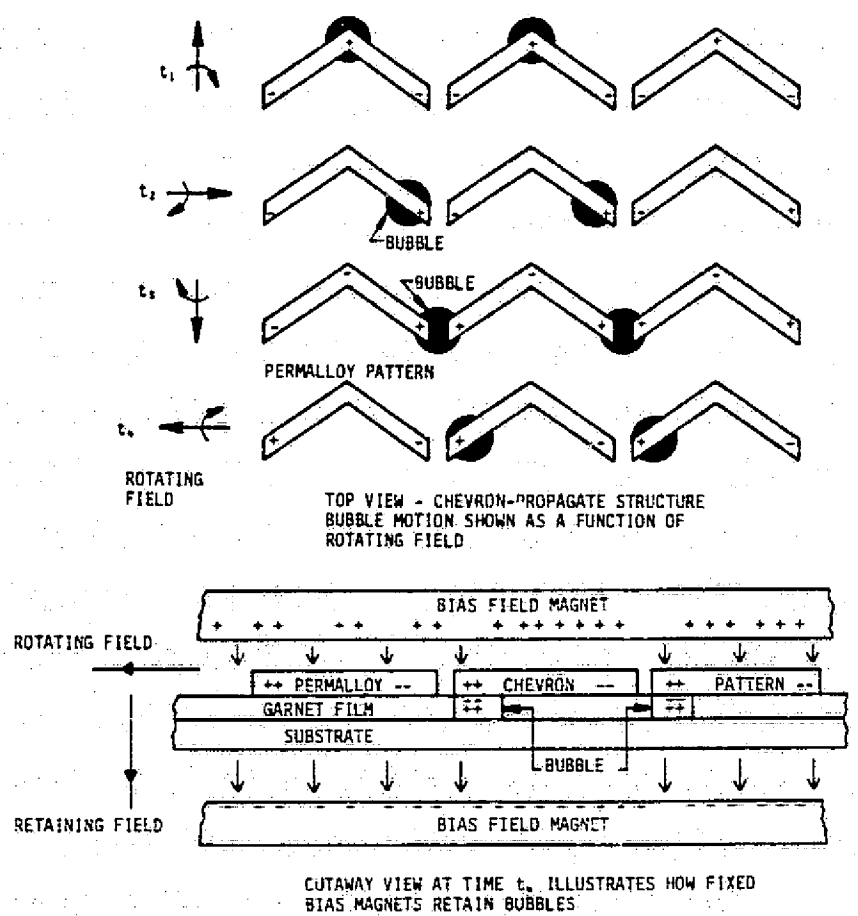


FIGURE 3.2-1. FIELD ACCESS BUBBLE PROPAGATION TECHNIQUE

3.2.1 State of the Art in Magnetic Bubble Memories

Magnetic bubble memory chips were introduced commercially for the first time during 1977. The first memory chip to enter the commercial market was the Texas Instruments TBM 101 chip. The chip, which had a storage capacity of 92 kbits, included two orthogonal drive coils to provide the rotating magnetic field, a permanent magnetic set, and a magnetic shield to protect the data from external fields. It was packaged in a 14-pin dual in-line module measuring 1.0 by 1.1 by 0.4 in. The chip architecture is major loop/minor loop, as illustrated in Figure 3.2.1-1, and contains a total of 157 minor loops each consisting of 641 bubble positions, resulting in a total capacity of 100,637 bits. To enhance production yields, as many as 13 of the 157 minor loops on the chip may be defective and the chip still meet specifications. The minimum data capacity of the remaining 144 good loops is therefore 92,304 bits.

In addition to the TI chip, another magnetic bubble system that saw limited commercial application during 1977 was a voice anncunce-ment system placed into service by Western Electric/Bell Laboratories. The 272-kbit system used four 68-kbit chips manufactured by Western Electric and was capable of storing a 12-sec voice message for use on the Bell telephone network.

The state of the art is to the point where development prototype and limited production chips in the 65- to 100-kbit range are fairly widespread (Ref. 3-2). Pilot production devices include a 262-kbit chip from Texas Instruments and a Western Electric/Bell Laboratories 272-kbit chip that is scheduled to replace the four-chip configuration in the voice announcement system identified above. Recently reported laboratory developments include 1-Mbit chips by both Rockwell and Bell Laboratories. The Rockwell 1-Mbit chip is fabricated on a 400- by 400-mil die with 1.8-μm bubbles and an 8-μm period. Other companies currently involved in bubble memory development work include IBM, Intel, National Semiconductor, Hewlett-Packard, Univac, Hitachi, Fujitsu, Plessey, and Phillips.

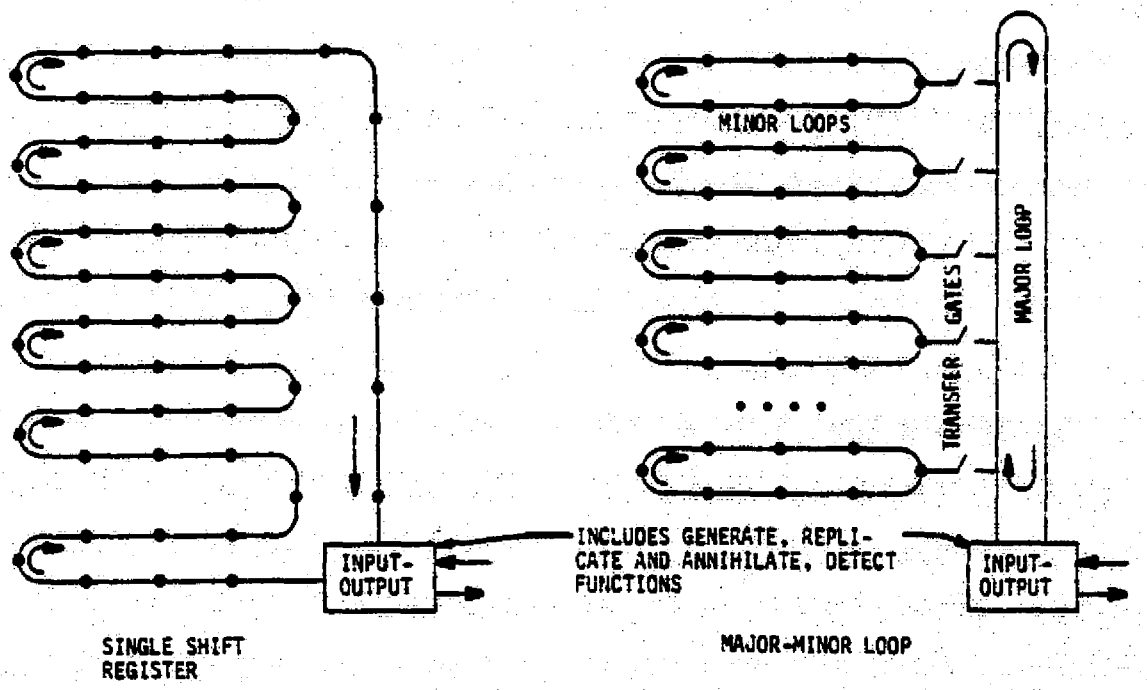


FIGURE 3.2.1-1. BASIC BUBBLE MEMORY ORGANIZATIONS

Increases in chip capacity are a major goal for designers of bubble memory systems. Improvements in chip capacity are accomplished through reductions in the bubble geometry, including the propagate structure. Figure 3.2.1-2 illustrates the four most prominent propagate structures that have been built to date.

Most bubble memory devices that are currently in production, including the 100-kbit Rockwell chip and the 92-kbit TI chip, have utilized the T-bar propagate structure. The T-bar, with two gaps per propagate period, has been recently replaced in most laboratories by more efficient single-gap structures such as the other three structures shown in the reference figure. The asymmetric half disk, which was invented at Bell Laboratories, provides a fourfold increase in density over the T-bar using the same wafer processing capability. Rockwell has utilized the asymmetric disk structure with 1.8-μm bubbles and an

1. T-BAR
   - 1/16 PERIOD MINIMUM FEATURES
   - DEFECT SENSITIVE BARS
2. ASYMMETRIC HALF DISK
   - 1/8 PERIOD MINIMUM FEATURES
   - NO BARS
   - NO CHANNEL-TO-CHANNEL INTERCONNECTS
3. ASYMMETRIC CHEVRON
   - GAP TOLERANT
   - INTERMESHED ADJACENT CHANNELS
4. CONTIGUOUS DISK
   - NO GAPS
   - IMPLANTED PROPAGATE STRUCTURE
   - 1/2 PERIOD MINIMUM FEATURES

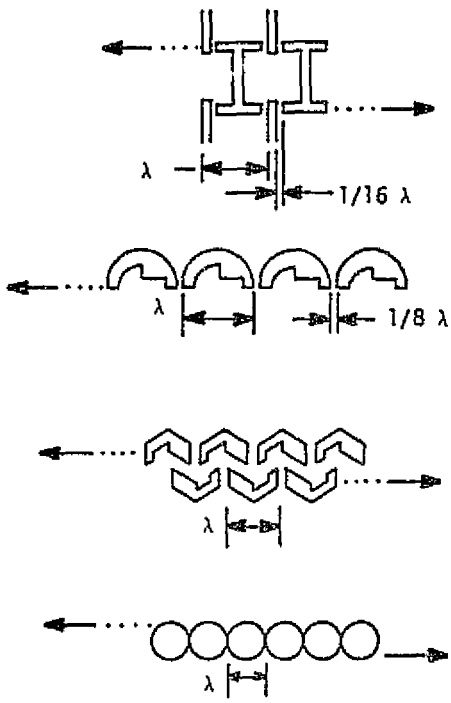


FIGURE 3.2.1-2. EVOLUTION OF PROPAGATE STRUCTURES

8-μm period to construct a 1-Mbit chip with a storage density of 1.6×10^6 bits/cm^2. The asymmetric chevron, which was also invented at Bell Laboratories, provides almost a factor of two improvements in chip density over the asymmetric half disk. The asymmetric chevron structure enables chip densities approaching 3 Mbits/cm^2 using current wafer-processing techniques. The contiguous disk, discussed in Section 3.2.3, is an experimental propagate structure that provides a factor of at least four increase in density over the asymmetric half disk.

At present there are two popular chip architectures for organizing the bubble shift register structures. The simplest organization is a single-loop shift register. A 100-kbit single-loop chip operated at a 100-kHz shift rate has an access time of 0.5 sec. For applications requiring improved performance, the major/minor-loop organization illustrated in Figure 3.2.1-1 is most frequently used.

The access time for the 100-kbit chip is reduced to less than 4 msec if a major/minor-loop organization is used. This improvement in performance is obtained with an attendant increase in bubble chip complexity, interface electronics, and control logic. Since total access time is a combination of the minor-loop access time plus the major-loop access time, replacing the major loop with a decoder circuit further decreases access time. Rockwell has reported success (Ref. 3-3) in decreasing access time through the use of hybrid systems involving a combination of major/minor loops and decoder organizations. Although decoders decrease access time, they are considered complex, they dissipate relatively high power, and a reliable decoder is difficult to achieve. Near-future applications for bubble memories as onboard spacecraft storage devices are for high-reliability tape recorder replacements; hence access time is not the major design consideration. The 10^8-bit bubble memory recorder currently being developed for NASA by Rockwell has an access time of 0.5 sec.

Data transfer rates for a bubble memory system are a function of drive frequency, chip density, chip organization, and system organization. Typical drive frequencies range from 50 Hz to 300 kHz, with several companies reporting chips capable of operation to 500 kHz. The maximum drive frequency is limited by the maximum bubble velocity and the repetitive circuit period, λ. For a bubble memory with an 8-μm period fabricated from garnet film with a maximum bubble velocity of approximately 1,000 cm/sec, the practical upper frequency limit is approximately 600 kHz. In systems such as the 10^8-bit NASA bubble memory, data rates much higher than the chip drive frequency are achieved by parallel reading or writing of multiple chips within the memory system.

Chip fabrication yields on the order of 20 to 30% have been reported. The application of improved propagate structures such as the asymmetric chevron proposed by Bell Laboratories with only one permalloy feature and only one gap per propagate period should further improve chip yield by lowering the statistical probability of processing defects that cause problems.

At least one memory system employing magnetic bubbles for space data storage is currently under development, and others are known to be in the study phase. Rockwell International's Autonetics Group is currently developing a 10^8-bit bubble memory system for NASA/Langley. The program is intended to determine the applicability of bubble technology to spacecraft data recording applications by design, development fabrication, and test of a flight-qualifiable, solid-state data recorder. Table 3.2.1-1 presents the specifications for this system and compares it with the NASA 10^{-8} and 10^{-9} bit tape recorders that are also under development.

TABLE 3.2.1-1. COMPARISON OF NASA 10^9-BIT SOLID-STATE RECORDER WITH STANDARD 10^8-BIT AND STANDARD 10^9-BIT MAGNETIC TAPE RECORDERS

	NASA 10^8-BIT BUBBLE MEMORY	NASA 10^8-BIT STANDARD RECORDER	NASA 10^9-BIT STANDARD RECORDER
Manufacturer	Rockwell	RCA	Odetics
Year Operational	1978	1976	1978
Volume	616 in^3	363 in^3	1,648 in^3
Weight	40 lb	14 lb	TBD
Storage Capacity	1×10^8 bits	4.5×10^8 bits	4×10^9 bits
Data Rate	1.2 Mbps (serial) 2.4 Mbps (parallel)	2 kbps to 2.56 Mbps	6.25 kbps to 8.19 Mbps
Power: Min/Max Standby	13.5 W/103 W TBD	7 W/17 W 2.3 W	18 W/46 W 1 W
BER	1×10^{-8}	1×10^{-6}(EOL)	1×10^{-6}(EOL)
MTBF	40,000 hr	30,000 hr	TBD

3.2.2 Trends in Bubble Memories

Research in bubble memories is proceeding on all fronts, including materials, lithography, propagating structure design, chip organization and configuration, fabrication processes, and packaging. The results are enabling improvements in a number of important areas, such as storage density and capacity per chip, higher yields on larger chips, improved temperature and bit error performance, improved packaging that enables bubble chips to be used with integrated circuit chips, improved fabrication techniques leading to one-step fabrication, lower cost, and increases in potential applications.

Currently, bubble memory chip designs are evolving in two directions, corresponding to identified market areas. One chip design is configured for maximum storage density and is intended for applications where access time is not critical, such as airborne tape recorder replacement. The other chip design is intended for disk replacement applications that require access times less than 1.0 msec. Figure 3.2.2-1 is a projection of the trend in bubble memory chip storage capacities for these two market areas through 1985. Chip capacities are currently increasing by a factor of four per year. This trend should continue through 1980. By 1985, commercially available chips should have approximately 1,000 times the storage capacity of currently available chips. However, smaller-capacity chips (2 to 8 Mbits) with faster access times will most likely dominate the bubble memory market. The larger-capacity bubble memory chip designs will incorporate on-the-chip, fault-tolerant circuit arrangements to maximize the processing yield for the large-capacity devices.

The use of electron-beam fabrication and X-ray lithography will enable chips with much larger capacities to be fabricated by increasing both storage density and chip size. Chips fabricated using photolithography are currently limited in size to approximately 1 cm^2 with 1-μm minimum features. Using electron beam fabrication and X-ray lithography, as discussed in Section 1, the resolution limit can be reduced to 0.2 μm or less. Electron-beam fabrication also makes possible much larger chips; however, it is not believed that chips significantly larger than 1 cm^2 will be economically practical.

The cost of commercially available bubble memory systems is currently about 50 millicents per bit, as shown in Figure 3.2.2-2. As seen in the figure, the system cost should drop to 1 to 3 millicents per bit by 1985, depending on memory size, access times, etc. These cost trends are based on the opinions of various experts for fast-access bubble memories as represented by curves A and B of Figure 3.2.2-1.

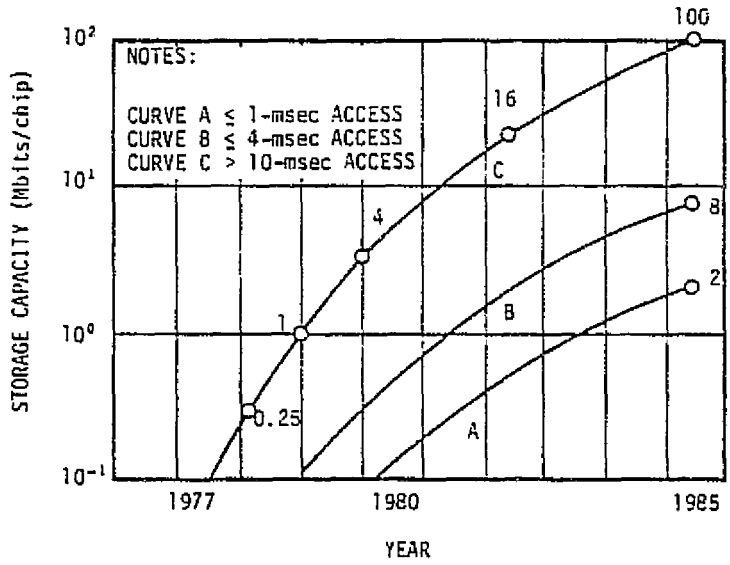


FIGURE 3.2.2-1. STORAGE CAPACITY PROJECTIONS FOR PRODUCTION BUBBLE MEMORY CHIPS

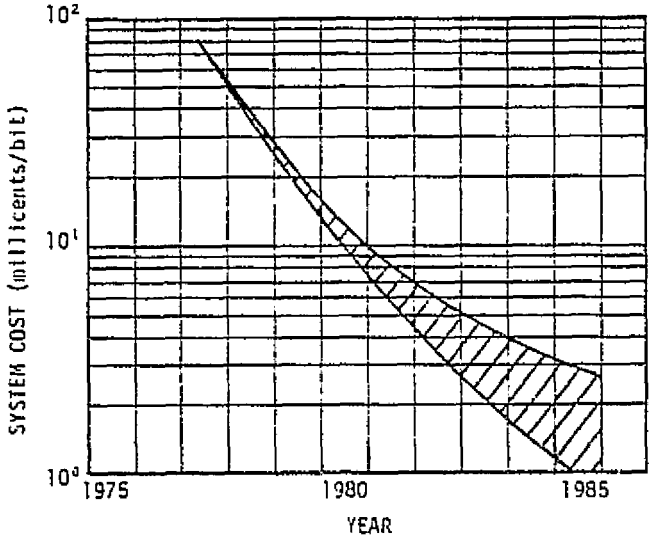


FIGURE 3.2.2-2. BUBBLE MEMORY SYSTEM COST PROJECTION (COMMERCIAL)

3.2.3 Projected Developments in Bubble Memories

The rapidly changing state of the technology in magnetic bubbles assures significant improvements over the existing state of the art. The trends presented in the previous section were based on trends in bubble technology in the 1973-1977 timeframe and on an assessment of the statements of a number of authorities in the field. The trends identified in Section 3.2.2 will result in the introduction of a number of products employing magnetic bubbles during 1978, with substantial increases in such products during the coming years. Applications that will employ bubble memories extensively by 1980 include microprocessors, minicomputers, smart terminals, desktop computers, point-of-sale terminals, and replacements for fixed-head disks and drums. Other potential applications for 1980 include military systems, aircraft and spacecraft recorders, data collection terminals, and special signal-processing devices. Applications projected for 1985 include all the above applications on a larger scale, including replacements for head-per-track disk and mass storage systems in general.

Several authorities have projected that chips with storage capacities of 10^8 bits/chip will be commercially available by 1985. This will be achieved by a combination of fabricating larger chips, improved propagate structures, and utilization of smaller device geometries. For example, the contiguous disk and bubble lattice device structures will provide 15 to 20 times the storage density of the T-bar structure used in the TI 92-kbit chip, and circuit fabrication resolution should increase by a factor of 5 to 10 times. Therefore, 10^8 bits/chip may be possible without increasing the chip die size.

The contiguous disk currently appears to offer the most promise as the next-generation bubble memory. The contiguous disk device uses propagating structures consisting of connected elements shaped like disks or diamonds (see Figure 3.2.1-2). Because of the absence of gaps that have to be resolved photographically, the density can be increased by at least a factor of four over the density of the asymmetric half disk.

The bubble lattice appears to offer a highly promising approach for long-range bubble systems. The bubble lattice is a fundamentally different type of bubble memory device from those that are currently available. Bubble lattice devices pack bubbles closely together in a hexagonal lattice and store the information in two kinds of bubbles with different magnetization. According to Ho and Chang of IBM's T. J. Watson Research Center (Ref. 3-4), the bubble lattice will necessitate more complicated structures than discrete bubble devices when implemented in functionally complete chips. IBM reported the development of an experimental 1-k bubble lattice chip at the 1977 Conference on Magnetism and Magnetic Materials (Ref. 3-5). The experimental device contains a 32-row by 32-column hexagonal array (like a honeycomb). The bubbles have a diameter of 5 μm and a center-to-center spacing of 11.5 μm, for a storage density of 5 Mbits/in^2 (0.78×10^6 bits/cm^2). Bubble lattice devices pack the bubbles much closer than conventional bubble devices; hence they have the potential for storing much more data in a given area than conventional bubble devices. For example, if the current IBM lattice device were built using the same bubble diameters as the Rockwell 1-Mbit memory, the lattice file would provide almost five times the storage density as the asymmetric half-disk structure. Future lattice devices should improve the storage density by an even greater factor by further reduction in the lattice bubble-to-bubble spacing.

In the bubble lattice device, every location in the lattice must always contain a bubble; hence information cannot be represented as the presense or absence of a bubble as in conventional bubble memory devices. Information is represented in the bubble lattice by the magnetic state of the "domain wall" that surrounds each bubble. The bubble structure can be forced into either of two magnetic states to represent "ones" and "zeros". Whenever a bubble is shifted out of the lattice, another one with the same information content is entered at the rear of the column to prevent the lattice from collapsing. The information stored in a particular bubble is read by denoting whether it moves forward along the line of the column or at a 30-deg angle from the line when it is removed from the lattice.

In addition to increased storage density, another advantage of the bubble lattice is that the bubble motion is controlled by electrical currents flowing through microcircuit patterns that are deposited directly on the electrically insulating bubble film material (yttrium iron garnet). This method of controlling bubble motion eliminates the rotating magnetic field and associated coils used in conventional bubble devices.

The major factors ultimately limiting high-density bubble chip designs are: high-resolution fabrication (line resolution), bubble-to-bubble spacing restrictions, and magnetic properties of small bubbles. Figure 3.2.3-1 illustrates how the aerial density changes as a function of minimum line width resolution and bubble size. If the line width is the limiting factor, the bubble lattice file and contiguous disk are about equal in bit packing density; however, if the bubble diameter is the limiting factor, the density of the bubble lattice file is much better than either the contiguous disk type or the types of bubble memory devices that employ permalloy propagate structures.

Although most authorities believe that bubble applications will be limited to storage requirements of 10^9 bits or less, Rockwell currently has an Air Force contract to study the feasibility of building an airborne 10^{12}-bit bubble memory system. Rockwell claims that this system could be operational before 1985 and the system goals represent a reasonable estimate of projected spaceborne bubble memory technology for that time period. The conclusion after completing the first year of the study is that most of the system design goals, with the exception of volume and power, can be achieved with current bubble technology, and it is reasonable to assume that a system that meets all the system goals can be built in the early 1980's. Table 3.2.3-1 compares the system goals for the Air Force 10^{12}-bit bubble memory system and NASA 10^8-bit bubble memory data recorder currently under development.

NASA's Langley Research Center is also developing a standard 10^7-bit bubble memory (SBM) for relatively low-storage-capacity applications. The SBM will weigh 15 lb and occupy a volume of 275 in^3. The SBM will have the capability for recording or playback of asynchronous data from dc to 1,024 Mbits/sec. The SBM will also provide the capability to simultaneously record and play back.

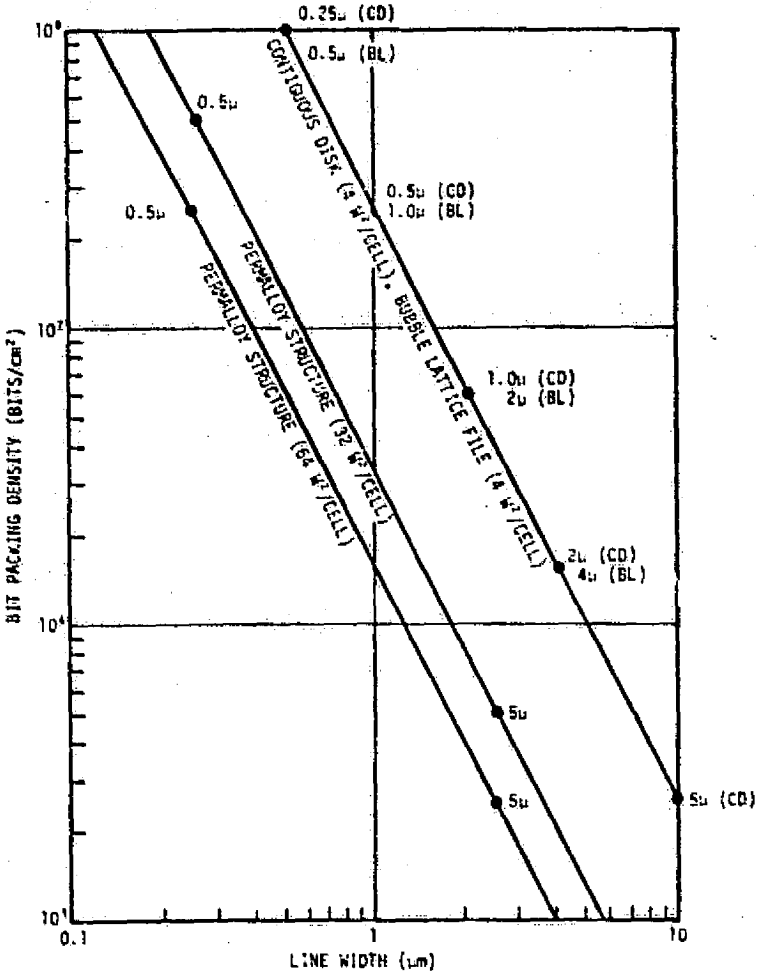


FIGURE 3.2.3-1. BIT PACKING DENSITY VERSUS LINE WIDTH AND BUBBLE SIZE FOR VARIOUS BUBBLE STRUCTURES

TABLE 3.2.3-1. COMPARISON OF NASA 10^8-BIT BUBBLE MEMORY DATA RECORDER AND AIR FORCE 10^{12}-BIT MEMORY SYSTEM GOALS

	NASA 10^8-BIT BUBBLE MEMORY	AIR FORCE 10^{12}-BIT BUBBLE MEMORY SYSTEM GOALS
Year Operational	1978	1985 (est.)
Volume	616 in^3	3,456 in^3
Weight	40 lb	120 lb
Storage Capacity	10^8 bits	10^{12} bits
Power (maximum)	103 W	200 W
Data Rate (maximum)	2.4 MHz	50 MHz

3.3 ELE(RON-BEAM ADDRESSABLE MEMORIES

Ele ron-beam addressable memories were considered as a potential space data : orage medium because of the high storage densities and rapid throu put rates projected for this technology. After evaluation, E-beams wer eliminated from contention as a possible storage medium for space for t following reasons:

- the time of this report, only one company icro-bit) is known to be actively pursuing beam memory system development, and all of eir research is being directed toward the mmercial market.

- e Micro-bit 950 system, which is scheduled r first delivery in 1979, is packaged in a binet that is 8 ft long by 5 ft high by 3 ft ep. The 950 system as physically packaged r the commercial market contains approxi- tely 2.1 Mbytes/ft^3, whereas the first- neration NASA/Rockwell 10^8-bit Solid-State corder (bubble memory) packages 35 Mbytes/ft^3 d the NASA/RCA 240 MBS tape recorder packages 750 Mbytes/ft^3.

- e economics of current E-beam system designs based on sharing the electronics among veral storage tubes. Therefore, the cost/ rformance of smaller systems (such as might packaged for space applications) that contain ly a few storage tubes is not attractive.

- ere are no near-future requirements for large ace data storage elements with capacity/access aracteristics that are uniquely characteristic beam systems.

Tech ically, an E-beam system could be packaged for space and should perfc n extremely well in the vibration-free space environment if sufficier justification existed to pursue development of the system, but this doe not currently appear to be the case. Early storage tube reliability roblems that were characteristic of small-area MOS targets have been sc ewhat overcome by use of larger storage tubes. Current estimates of storage tube reliability is about 5 years for both the cathode and ne target. A detailed discussion of commercial E-beam technology i presented in Section 7.2.2.1.

REFERENCES - SECTION 3

3-1 A. H. Bobeck, "The Development of Bubble Memory Devices", Electro 77, New York, New York

3-2 E. A. Torrero, "Bubbles Rise from the Lab", IEEE Spectrum, September 1976

3-3 T. T. Chen, T. R. Oeffinger, and I. S. Gergis, "A hybrid Decoder Bubble Memory Organization", IEEE Transactions on Magnetics, Vol. Mag-12, No. 6, pp. 680-682, November 1976

3-4 H. Chang and C. P. Ho, "Field-Access Bubble-Lattice Propagation Devices, IEEE Transactions on Magnetics, Vol. Mag-13, No. 2, March 1977

3-5 "IBM Researchers Testing 1K Bubble Lattice", Computer World, p. 64, December 12, 1977

3-6 J. M. Hayes and F. D. Dell, Wideband Image Recorder for the Earth Resources Technology Satellite", 1971 National Telemetering Conference, April 12-15, 1971

3-7 A. M. Burke and F. L. Putzrath, "NASA Standard Spacecraft Recorder", Recording Systems, RCA Government Communications Systems, Camden, New Jersey

3-8 J. D. Rittenhouse, "Rotary Head Instrumentation Recorders for Telemetry Recording", Telemetry Journal, Vol. 3, No. 6, October/November 1968

3-9 J. C. Mallinson, "Trends in High Rate High Density Digital Recording", Countermeasures, pp. 99-101, May 1977

3-10 O. E. Bessette, "Recent Advances in Magnetic Recording of Digital Data", International Telemetry Conference, Los Angeles, California, September 28-30, 1976

3-11 M. Wildmann, "Mechanical Limitations in Magnetic Recording", IEEE Intermag Conference, Toronto, Canada, May 14-17, 1974

3-12 R. G. Kiwimag, "Channel Coding for Digital Recording", IEEE Intermag Conference, Toronto, Canada, May 14-17, 1974

3-13 Odetics, Inc., "Odetics Spaceborne Recorders", 1859 S. Manchester Avenue, Anaheim, California, November 16, 1976

3-14 C. R. Thompson and J. L. Waring, "Digital Data Recording at Five Million Bits per Square Inch", International Telemetering Conference, 1975

3-15 P. I. Bonyhard and J. L. Smith, "68 Kbit Capacity 16 μm-Period Magnetic Bubble Memory Chip Design with 2 μm Minimum Features", IEEE Transactions on Magnetics, Vol. Mag-12, No. 6, pp. 614-617, November 1976

3-16 J. M. Salzer, "Bubble Memories - Where Do We Stand?", Computer, pp. 36-41, March 1976

3-17 J. E. Geusic, "Bubble Memory Design and Performance", IEEE Transactions on Magnetics, Vol. Mag-12, No. 6, p. 622, November 1976

3-18 N. Saito, T. Toyooka, S. Yoshizawa, and Y. Sugita, "A High Speed Drive Coil for a Large Capacity Bubble Memory", IEEE Transactions on Magnetics, Vol. Mag-12, No. 6, pp. 639-641, November 1976

3-19 R. F. Baily and J. E. Ypma, "Bubble Domain Memories", Solid State Technology, pp. 74-82, September 1976

3-20 M. S. Cohen and H. Chang, "The Frontiers of Magnetic Bubble Technology", Proceedings of the IEEE, Vol. 63, No. 8, pp. 1196-1206, August 1975

3-21 A. H. Bobeck, P. I. Bonyhard, and J. E. Geusic, "Magnetic Bubbles - An Emerging New Memory Technology", Proceedings of the IEEE, Vol. 63, No. 8, pp. 1176-1195, August 1975

3-22 M. Takasu, H. Maegawa, S. Furuichi, M. Okada, and K. Yamagishi, "A Fast Access Memory Design Using 3 μm Bubble 80K Chip", IEEE Transactions on Magnetics, Vol. Mag-12, No. 6, pp. 633-635, November 1976

3-23 C. D. Mee, "A Comparison of Bubble and Disk Storage Technologies", IEEE Transactions on Magnetics, Vol. Mag-12, No. 1, January 1976

3-24 A. S. Hoagland, "Magnetic Recording Storage", IEEE Transactions on Computers, Vol. C-25, No. 12, pp. 1283-1288, December 1976

3-25 P. Uber, "1×10^9 Bit Standard Spacecraft Tape Recorder", NASA No. 73-15033, January 1974

3-26 O. E. Bessette, "High Capacity, High Data Rate Instrumentation Tape Recorder System", International Telemetering Conference, October 9-11, 1973

3-27 T. T. Chen and B. A. Waggoner, "10^{12} Bit Bubble Memory System", NAECON, Dayton, Ohio, May 17-19, 1977

3-28 J. E. Juliussen, "Magnetic Bubble Systems Approach Practical Use", Computer Design, pp. 81-91, October 1976

3-29 A. H. Bobeck and I. Danylchuck, "Characteristics and Test Results for a 272K Bubble Memory Package", 1977 IEEE Intermay Conference, Los Angeles, California

4. SPACE DATA HANDLING ELEMENTS

Space data handling as used herein encompasses all functions that occur between the output of the data source (sensors, computers, imaging systems, etc.) and the input to the downlink modulator. The standard functions of signal conditioning, multiplexing, analog-to-digital and digital-to-analog conversion, intra-spacecraft communications, data formatting, and data buffering are included, as well as those functions that are derivatives of the standard functions, such as onboard data base management, data compression, error correction procedures, and adaptive systems. Figure 4-1 organizes and depicts the standard data handling elements for digital data systems, and Figure 4-2 presents the corresponding information for analog data systems. The major emphasis throughout the text is on digital systems, since the trend for future data systems is heavily oriented in that direction.

Space data handling systems offer the greatest potential of any single data system element for improving the performance and cost-effectiveness of future NASA data systems. Technology advances in data handling will result from improvements in device speed, storage capacity, size, power consumption, and cost; but the most significant advances will be the result of the implementation of data handling system architectures that have the capability to select, process, format, and control the flow of information to the ground. Included in this category are intelligent sensors; asynchronous data collection systems; high-speed processors; high-capacity, fast throughput buffers; and packet-handling communication links.

This section discusses the devices that are unique to space data handling, such as signal conditioners, A/D and D/A converters, and multiplexers. These devices form only a portion of the data handling system and must be considered in conjunction with technology in other areas, such as space processors, magnetic bubble memories, semiconductor memories, microprocessors, and mass storage system. The systems' impacts of interconnecting all of these devices are presented in Section 1.

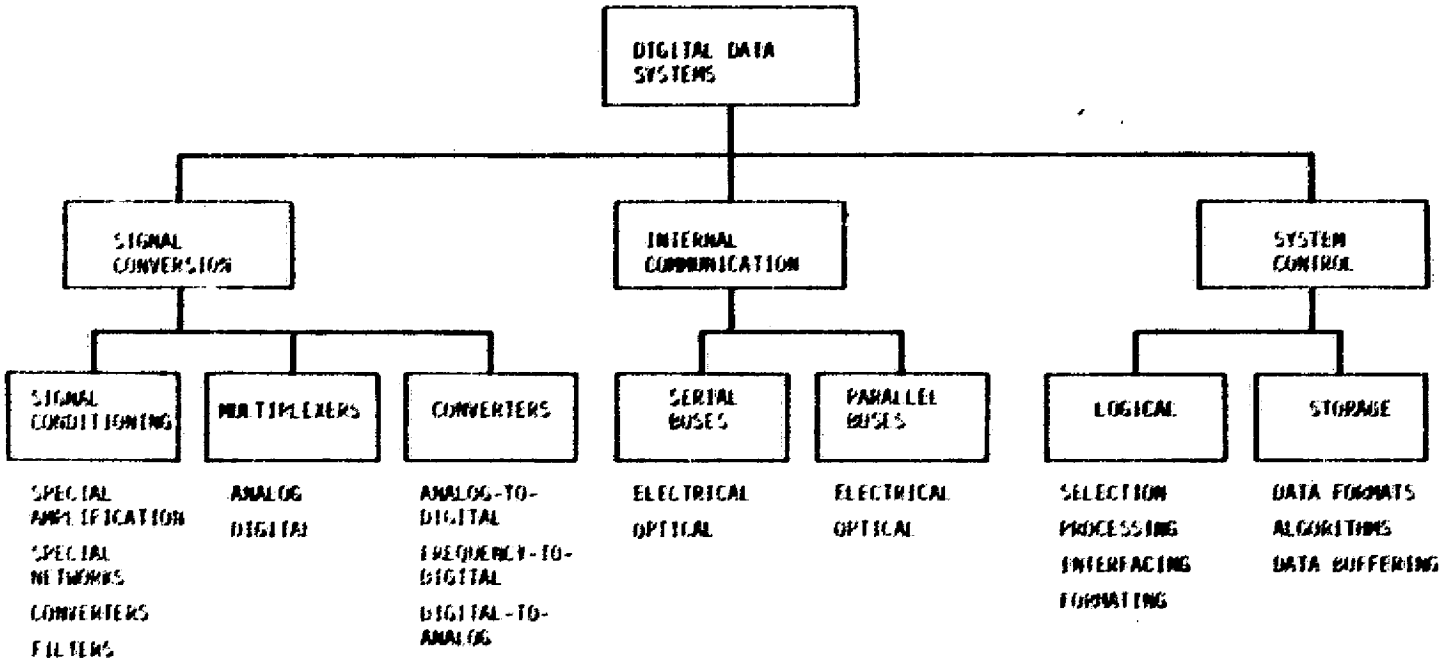


FIGURE 4-1. DIGITAL DATA SYSTEM ELEMENTS - DATA HANDLING

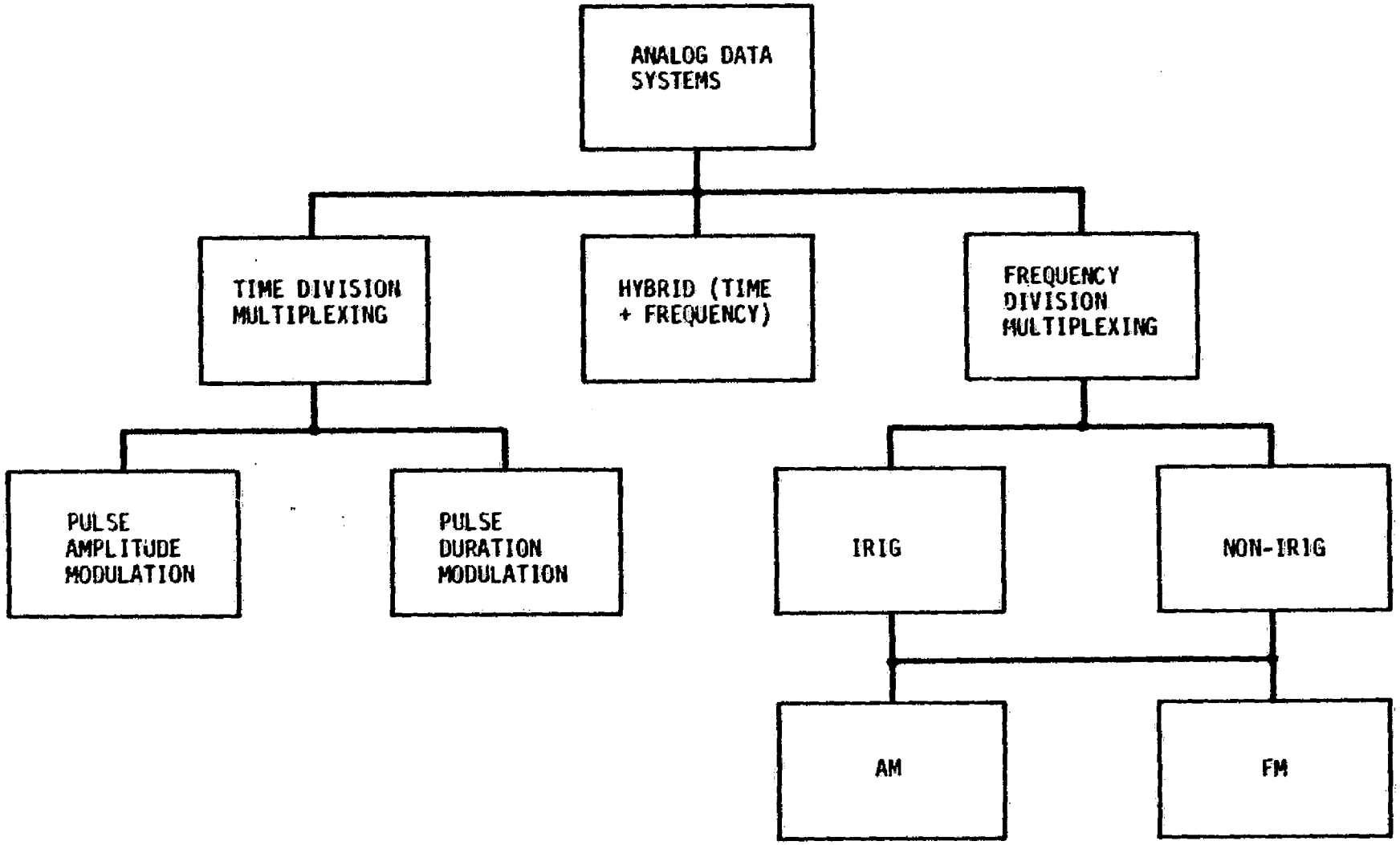


FIGURE 4-2. ANALOG DATA SYSTEM ELEMENTS - DATA HANDLING

4.1 DIGITAL DATA HANDLING SYSTEMS

Digital Data Handling Systems (DDHS) technology is discussed in subsequent paragraphs in terms of signal conversion elements, intra-spacecraft communications, and system control. Element configurations and the functions typically associated with these elements are depicted in Figure 4.1-1. Digital data handling systems included within the categories discussed herein may serve as an input to a communications downlink and/or as an onboard data management system with the capability to adapt to current conditions. Technology advances will make onboard data management more technically and economically feasible and thus more prevalent on future NASA missions, both manned and unmanned.

Technology aspects of signal conversion and intra-spacecraft communications are discussed at both the component and the system level within this section, and DDHS control is discussed at the system/functional level. Component-level elements of system control are covered in other sections. Sections 2 and 7 present processor technology; Section 3 presents space data storage device technology, including magnetic bubbles; and Section 7 covers semiconductor memories, including charge-coupled devices. Section 1 addresses system-level considerations in general, with specific applications to digital data handling, which is one of the more important elements of data systems from the standpoint of future improvements.

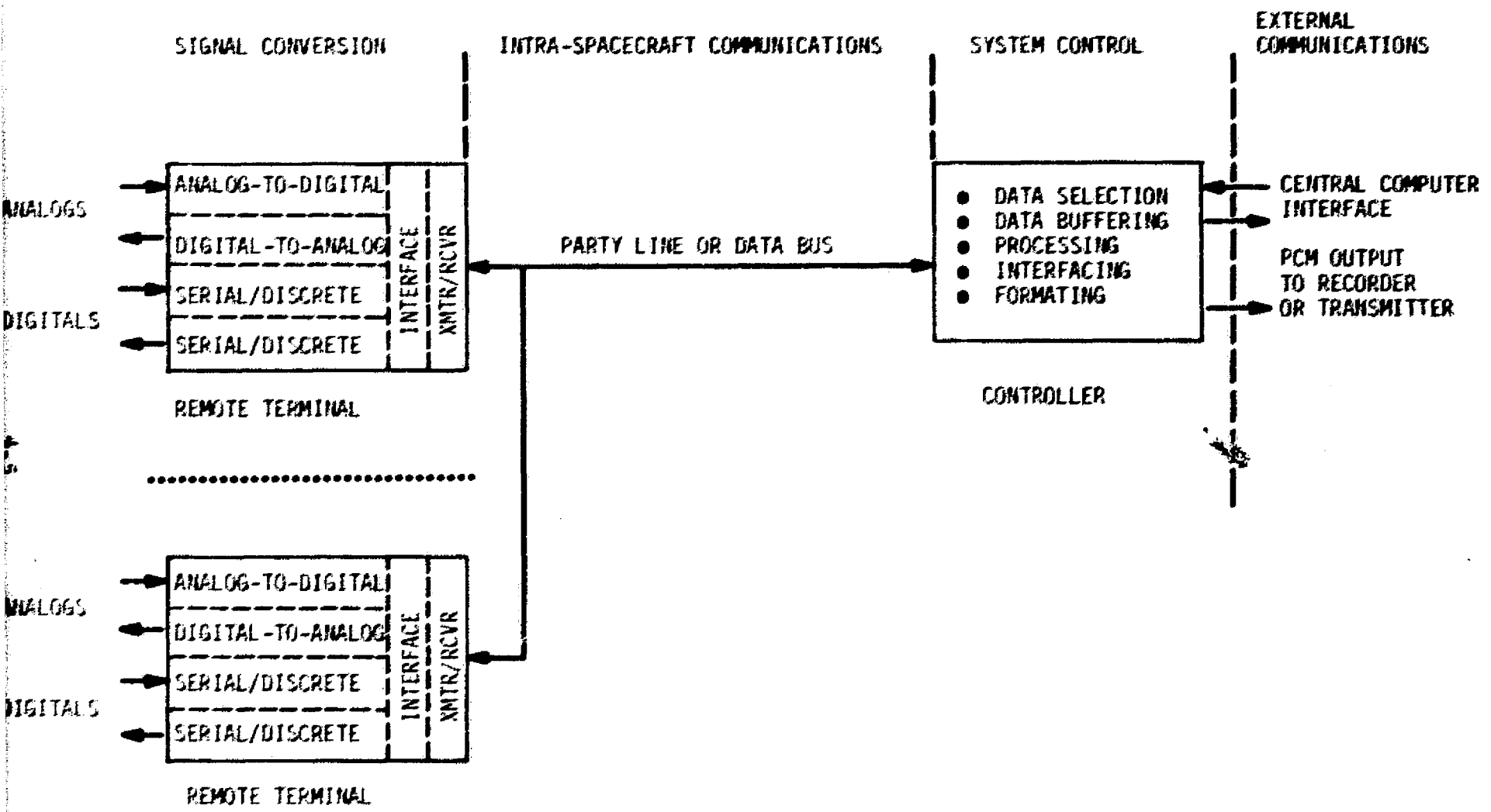


FIGURE 4.1-1. MAJOR FUNCTIONAL ELEMENTS OF DIGITAL DATA HANDLING SYSTEMS

4.1.1 State of the Art in Digital Data Handling Systems

Digital data handling systems may be classified into two basic categories: namely, general-purpose and unique/dedicated systems. Although this classification is not particularly significant to technology developments, the component technology that is applied to each of these system types is different as the result of the different requirements that characterize these two different types of data handling systems. Table 4.1.1-1 presents the general characteristics for the two categories. The principal differences between the two categories are data rate and flexibility. Whereas the general-purpose systems typically support lower data rates, the requirements for flexibility are generally more stringent. Circuit technologies capable of providing the high speeds associated with the dedicated systems have typically consumed relatively high power and it has not been feasible to provide both the high data rates and the flexibility. Cost has also been a factor. Although the speed-power products of the newer circuit technologies are improving along with cost, the uniqueness of most high-data-rate systems tend to make it more feasible from both a cost and performance viewpoint to retain the basic characteristics illustrated in Table 4.1.1-1.

General-purpose digital data systems include the various types of low- to medium-data-rate PCM telemetry and instrumentation systems, ranging in size from small PCM multiplexer/encoders to relatively large, remote, multiplexed instrumentation systems that can accommodate several hundred input channels. Typical systems that are available for space applications operate at rates ranging from 25 Kbits/sec to 5 Mbits/sec. Some custom designs go as high as 10 Mbits/sec. Such systems are available from a number of companies, including: Aydin Vector, Base Ten Systems, Gulton, Spacetac, Sperry Phoenix, SCI Systems, and Teledyne Telemetry. These general-purpose systems are normally capable of accepting various combinations of high-level and low-level analog signals and discrete (bi-level) signals. Additionally, most offer some flexibility in selection of sample rates, word length, bit rate, and synchronization codes via both programmable and hardware options. Systems in this

TABLE 4.1.1-1. STATE-OF-THE-ART DATA SYSTEM CHARACTERISTICS

CHARACTERISTIC	GENERAL-PURPOSE DATA SYSTEM	UNIQUE-DEDICATED DATA SYSTEM
System Type	PCM Telemetry or Computer Interface	PCM-High Data Rate
Output Data Rate	<10 Mbits/sec	>5 Mbits/sec
I/O Format	Variable/Programmable	Fixed
IRIG-Compatible	Generally, Yes	Generally, No
Availability	Off-the-Shelf or Custom Design	Custom Design
Number of Inputs	30 to 1,000	1 to 10
Function-Typical	Payload Data/Scientific Data/Housekeeping Data	Visible Imaging/ IR Imaging/Digital Radar
Typical Hardware	Shuttle MDM/IUS Data System	Multispectral Scanner/ VISSR

classification are often available as off-the-shelf hardware. These systems normally provide PCM outputs that are IRIG-compatible. Not all general-purpose data systems are characterized by a PCM output to a data collection ground station. Some data systems, such as those used in advanced military aircraft, communicate via a data bus to a central computer that provides the data monitoring, display, and control functions in real time. The hardware used in general purpose aircraft data systems (such as the B-1 CITS) is very similar to data systems hardware used in space applications (such as the Space Shuttle MDM).

The unique/dedicated digital data systems include these systems that are either dedicated to a single high-data rate sensor or require unusual data-handling capabilities that are not practical to implement with a general purpose data system. Examples of unique/dedicated data systems are the various types of IR and visible imaging systems, such as the Multi-Spectral Scanner data system on Landsats 1 and 2 and the Thematic Mapper on Landsat-D.

Classification by application provides a convenient means for comparing technology at the system level; however, it does not provide a convenient subdivision for discussing component-level technology. Subsequent discussions of space data handling systems are therefore divided into the following functional elements:

- Signal Conversion
- Intra-spacecraft Communications
- Systems Control.

This subdivision provides a convenient means for subdividing and grouping the various types of circuit technology used in various data system designs. The interrelationships of these functional elements for a typical data system was illustrated in Figure 4.1-1. The following sections discuss the state of the art, trends, and projected developments in digital data systems from the standpoint of both system-level and component-level technology.

4.1.1.1 Signal Conversion - Signal conversion hardwar_ includes signal conditioners, analog/frequency-to-digital converters, digital-to-analog converters, and analog/digital multiplexers that are used to interface the data system with other spacecraft subsystems. Some of the more frequently encountered signal conversion hardware elements include:

- Signal conditioners
  - Special amplifiers
  - Special networks
  - Sample-and-hold circuits
  - Filters
- Multiplexers
  - High-speed analog
  - Precision analog
  - Digital
- Converters
  - Analog-to-digital
  - Frequency-to-digital
  - Digital-to-analog.

Signal conditioning circuitry is generally used to either transform voltage levels, isolate grounds, filter inputs, or provide analog signal sampling. In addition, signal conditioning circuits are also used to provide special functions such as transducer excitation/compensation or thermocouple reference compensation. In most data systems, the signal conditioning circuitry filters, scales, etc., the input signals so as to provide standard voltage levels to the multiplexer input.

The multiplexer allows a number of analog inputs to share a common analog-to-digital converter or to simply provide sample sequencing or selective buffering in the case of a digital multiplexer.

The converters are used to transform analog signals into a digital format that is compatible with the data system or to transform digital outputs into their analog form to provide spacecraft control signals.

4.1.1.1.1 Signal Conditioning - Data system designs have traditionally utilized signal conditioning circuits between the outputs of the various sensors or transducers and the inputs to the data system multiplexer. Functions performed by the signal conditioning circuitry include: scaling, filtering, offsetting, demodulation, buffering, isolation, and nonlinear amplification.

Many state-of-the-art data systems eliminate much of the conventional signal conditioning circuitry by the use of differential analog multiplexers with programmable gain/programmable offset amplifiers. This technique effectively places the multiplexer ahead of the signal conditioning circuitry, thus eliminating the need for providing separate signal conditioners for each individual channel. Programmable gain and programmable offset also increase system flexibility by allowing a single input channel to function either as a high-level channel or a low-level channel simply by reprogramming the format controller. Programmable gain ranges that are available in state-of-the-art systems range from less than one to greater than 500.

Signal conditioning technology advances have been closely tied to advances in linear circuit technology and advances in hybrid circuits, especially the ability to fabricate precision, stable, thin-and-thick-film resistor networks. Table 4.1.1.1-1 illustrates typical component types that are required to fabricate different types of signal conditioning circuitry.

In general, the degree to which LSI can be employed in linear circuitry is limited compared with digital technology because linear circuits must withstand higher voltages, typically 30 to 40 V. Therefore, oxides in linear devices must be 5 to 10 times thicker than for digital circuitry; hence the resulting device geometries are much larger. Also, certain components, such as precision capacitors used in precision integrators, are not available in a package suited for hybrid design. Therefore, these type components must be packaged outside the hybrid or monolithic circuit module or an alternative design approach must be used. One such alternative design approach is to use digital circuitry if possible to replace analog circuitry. For example, an analog integrator can be replaced by a voltage-controlled oscillator and a digital converter.

TABLE 4.1.1.1-1. EXAMPLE OF SIGNAL CONDITIONING CIRCUIT COMPONENT USAGE

SIGNAL CONDITIONING FUNCTIONS	ELECTRICAL COMPONENT TYPES										
	OPERATIONAL AMPLIFIER	HIGH INPUT Z OPERATIONAL AMPLIFIER	MOS TRANSISTOR	BIPOLAR TRANSISTOR	REFERENCE DIODE	CLAMP DIODE	COMPARATOR	PRECISION RESISTOR	STANDARD RESISTOR	CERAMIC CAPACITOR	PRECISION CAPACITOR
Demodulator, 400 Hz	X		X				X	X	X	X	X
Thermocouple	X		X					X	X	X	
Instrumentation Amplifier		X				X		X			
Presampling Filter	X					X		X		X	
Resistance Thermometer	X		X					X	X	X	
Bridge Network with Excitations	X			X				X	X	X	
Logrithmic Amplifier	X					X		X	X	X	
Signal Conditioning Assembly Power Supply	X			X	X	X		X	X	X	

Characteristics of state-of-the-art operational amplifiers are presented in Table 4.1.1.1-2. Two recent innovations in amplifier technology are BI-FET operational amplifiers, in which FET and bipolar devices are fabricated on a single chip, and thick-film hybrid isolation amplifiers. Prior to the recent Burr-Brown announcement, all commercially available isolation amplifiers were packaged as modules. The BI-FET amplifiers provide extremely high input impedance, excellant gain linearity, and low cost.

Hybrid circuit packaging techniques are used to achieve small package size in many state-of-the-art data system designs. The adaptation of hybrid packaging to signal conditioning applications is often limited because many of the specialized components used in signal conditioners (i.e., transformers) cannot be readily adapted to hybrid designs. According to various authorities, there have been no radical changes in hybrid circuit packaging technology during the last 5 years. The packaging of hybrid circuits has been evolving slowly during the past 5 years and should continue to do so in the immediate future.

4.1.1.1.2 Multiplexers - Space data system designs utilize a variety of analog and digital multiplexer configurations. Multiplexing as applied to signal conversion includes the analog and digital multiplexers used to interface with other spacecraft subsystems. Multiplexing functions not associated with signal conversion include data bus multiplexing and the interleaving of data from the various subsystems of the data system. These types of multiplexer functions are discussed elsewhere in Section 4.

Some of the most demanding state-of-the-art multiplexer requirements are associated with high-data-rate imaging sensors such as the Landsat-D Thematic Mapper. These type systems require analog multiplexer speeds on the order of 10 Msamples/sec. Only a few spaceborne multiplexers have been designed to operate at these high rates. One recently developed system was the "Multimegabit Operation Multiplexer System" (MOMS, Ref. 4-2), which was developed by Harris Corporation for the Goddard Space Flight Center. The system was developed to prove the feasibility of implementing

TABLE 4.1.1.1-2. COMPARISON OF STATE-OF-THE-ART AMPLIFIER CHARACTERISTICS

TECHNOLOGY	AMPLIFIER CLASSIFICATION	INPUT OFFSET VOLTAGE	INPUT BIAS CURRENT	UNITY GAIN BANDWIDTH (Hz)	SLEW RATE (V/μsec)	PRICE ($)
Monolithic	Low offset, low drift	50 μV to 0.5 mV	1 to 10 nA	400K to 1M	0.1	2
	Precision, low bias	0.5 to 6 mV	50 pA to 2nA	500K to 1M	0.1	2
	Chopper-stabilized	50 to 80 μV	150 pA	3M	2.5	25
	Bipolar input	1 to 5 mV	20 nA	100K to 1M	1 to 5	5
	BI-FET	15 mV	20 pA	100 K to 1M	1 to 5	5
Hybrid	Low Offset, Low drift	From 8 μV	From 10 pA	1.5M	0.6	10
	High voltage, high current	From 2 mV	From 20 pA	5M	30	50
	Bipolar input	0.2 mV	40 nA	2M	10	15
	FET input	5 mV	10 nA	600K	2	25
Modular	FET input	200 μV to 1 mV	0.1 to 25 pA	500K to 5M	0.25 to 1,000	50
	Precision chopper	10 to 25 μV	50 to 150 pA	100M	0.2 to 100	75
	Wideband	1 to 3 mV	100 pA to 5nA	60M to 120M	300 to 1,000	75
	Precision electrometer	1 to 10 mV	10 fA	2,000M	0.0004	100

a high-rate data system for the Landsat-D Thematic Mapper application. MOMS consists of two 140-Mbit/sec subsystems from which the outputs are interleaved to produce a combined system output data rate of 280 Mbits/sec. Each of the 140-Mbit/sec subsystems consists of a multiplexer with 30 radiometer channels and 2 vidicon channels, sample-and-hold circuits, and a 20-Msample/sec by 7-bit analog-to-digital converter. Although MOMS was only implemented as a breadboard, the system has a projected volume of 625 in^3 and a power requirement of 30 W. The MOMS multiplexer uses a two-level gating arrangement that employs a sophisticated bipolar analog switch with a 12-nsec settling time. The sample-and-hold circuits are used only for the two vidicon channels.

LSI multiplexer chips that include 16 differential analog channels with channel selection decoders have been available for several years; however, the use of LSI circuitry for spaceborne analog multiplexer applications is often limited by system design constraints that include requirements that do not allow a multiplexer circuit failure to affect multiple channels.

A number of commercial monolithic data conversion components have begun to emerge with the multiplexer and converter functions integrated into a single chip. These converter subsystem chips have been initially offered by National Semiconductor and Fairchild and provide an analog multiplexer with an 8-bit converter at about one-fifth the cost of similar hybrid devices. In fact, it is becoming increasingly difficult to distinguish between what is a component and what is a subsystem.

4.1.1.1.3 Converters - Numerous techniques are used to perform analog-to-digital conversion, including successive approximation, integrate and count schemes (such as double ramp and triple ramp), and high-speed parallel. Depending on the application and conversion technique employed, conversion speeds vary from 10 nsec to 3 sec and accuracies from 4 to 16 bits. Converters utilized in spacecraft applications generally require conversion rates that are faster than 50 μsec, and the following discussion is limited to this type converter.

The primary methods used to perform relatively high-speed analog-to-digital conversion are:

- Successive approximation
- Series-parallel
- Parallel
- Cyclic.

State-of-the-art characteristics of converters employing these techniques are presented in Table 4.1.1.1-3, which indicates a wide variation in the cost and performance of the different types of converters. This is due in part to the current rapid emergence of monolithic converters, which are providing a significant reduction in converter cost per unit of performance. Monolithic, successive-approximation, two-chip analog-to-digital converters are now available with resolutions up to 10 bits. Precision Monolithics and Harris Semiconductor have both recently announced single-chip, 12-bit digital-to-analog converters and Analog Devices recently announced the industry's first 10-bit, successive-approximation analog-to-digital converter that is completely self-contained on a single chip.

Successive approximation is the most popular method for medium- to high-speed (1 to 50 μsec) analog-to-digital conversion. Successive-approximation converters use either current summing or voltage summing ladder networks in which successively smaller voltages are sequentially summed and compared with the input or unknown voltage. After each comparison, the previously added voltage is either removed or retained depending on whether the sum is greater than or less than the unknown input as determined by the converter's voltage comparator. The speed of successive approximation converters is limited by the number of bits (successive operations) and by the gain-bandwidth products required to resolve each voltage comparison to the required accuracy. One major advancement that has been made in the design of successive approximation converters is the ability to circumvent transient and settling time problems that are inherent in these types of analog-to-digital converters by the addition of an error correction feedback technique. SCI Systems

TABLE 4.1.1.1-3. STATE-OF-THE-ART IN ANALOG-TO-DIGITAL CONVERTERS

TYPE	CONVERSION TIME	PRICE RANGE ($)	TECHNOLOGY	NOTES
General-Purpose				
8-bit	1 to 20 μsec	8 to 190	Monolithic/Hybrid	Successive Approximation
10-bit	2 to 30 μsec	25 to 330	Monolithic/Hybrid	Successive Approximation
12-bit	2 to 50 μsec	50 to 350	Hybrid/Module	Successive Approximation
High-Speed				
6-bit	10 nsec minimum	TBD	Monolithic/Hybrid	*
8-bit	20 nsec to 1 μsec	225 and up	Monolithic/Hybrid	*
10-bit	100 nsec to 2 μsec	250 and up	Monolithic/Hybrid	*
High-Accuracy				
14-bit	6 μsec minimum	420 and up	Module	Successive Approximation
16-bit	8 μsec minimum	795 and up	Module	Successive Approximation

*Various methods of high-speed A/D conversion are discussed in the text.

has developed a converter that has the ability to correct encoding errors prior to the final decision provided the net decision error is within certain bounds. This type of converter is useful in systems utilizing multiplexed programmable amplifiers with gains of up to 500 that require significant settling time. With a self-correcting converter, the conversion can begin before the data input has completely settled.

The analog-to-digital converter used in the MOMS, previously discussed, is a series-parallel type that consists of two series stages, each of which contains 16 parallel comparators. The first stage encodes the four most significant bits. These four bits are then converted to analog by a digital-to-analog converter and subtracted from the input signal. The second stage encodes the subtracted difference. Since completing the MOMS contract, Harris has redesigned the analog-to-digital converter to provide 8-bit resolution at 30 Msamples/sec.

In addition to the series-parallel type of analog-to-digital converter used in MOMS, two other types of ultra-high-speed analog-to-digital converter designs are frequently used in high-rate data systems. These are the cyclic converter and the parallel converter.

The parallel converter is a "brute force" approach the utilizes 2^{N-1} comparators to implement an N-bit converter. Although the parallel type converter is very fast (<20 nsec), except for use in low-resolution applications, most designs to date are too large and require too much power to be used in aerospace data systems.

The cyclic analog-to-digital converter consists of a series of nonlinear amplifier stages each with a transfer function as shown in Figure 4.1.1.1-1. The converter is named cyclic because the voltage at the outputs of each stage cycles up and down as the input voltage varies linearly. The cyclic converter has the advantage of high speed with minimum circuitry. The main disadvantage of the cyclic converter is accuracy, since accuracies in excess of 7 bits are difficult to achieve. The cyclic converter binary representation is inherently generated in Grey code. For applications requiring binary, grey-code-to-binary converters are used.

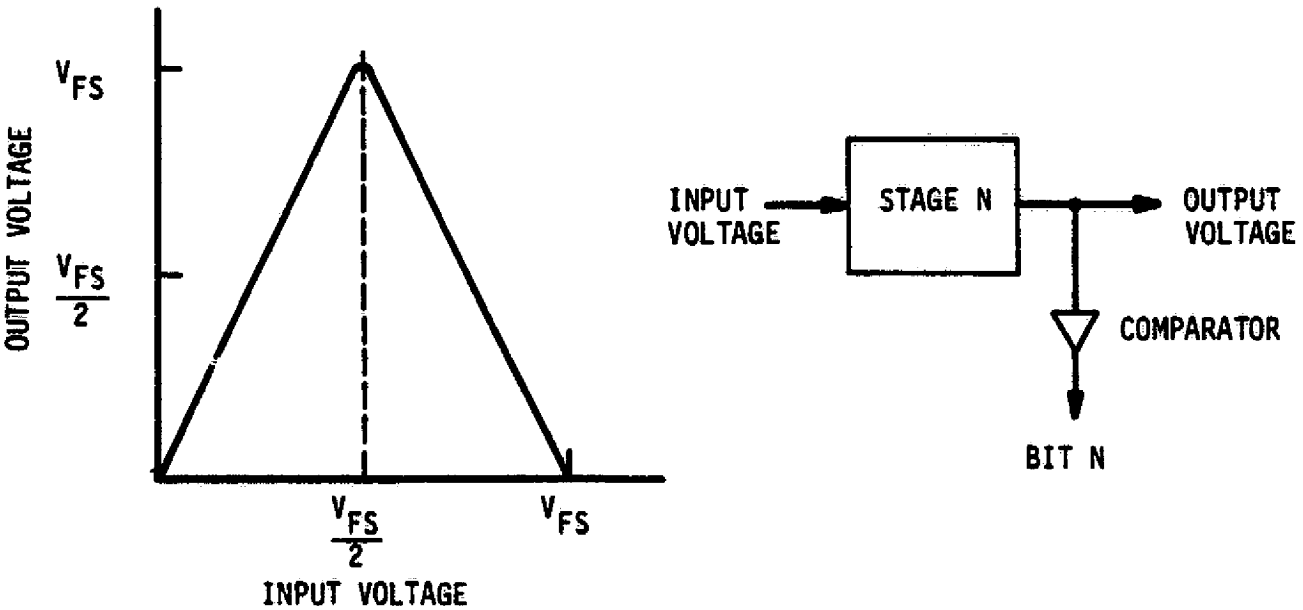


FIGURE 4.1.1.1-1. CYCLIC ANALOG-TO-DIGITAL CONVERTER STAGE TRANSFER FUNCTION

The Computer Labs MATV series of analog-to-digital converters overcomes the accuracy limitation of the cyclic converter by using cyclic scheme to encode the six most significant bits and a parallel encoder for the two least significant bits. The MATV converters provide 8-bit accuracy and are packaged in a configuration suitable for aerospace applications. Other specifications include a conversion time of 50 nsec, power consumption of 8 W, and size of 5.5 by 4.8 by 0.85 in.

Advanced Micro Devices, Inc., is currently developing a 4-bit quantizer chip with 8-bit accuracy. Each 4-bit quantizer chip contains 16 high-speed comparators with an associated resistor divider network and logic packaged as a single LSI integrated circuit. Using two of these devices plus some additional circuitry, a series-parallel 8-bit converter with a 20-Msample/sec word rate can be built. Sixteen of these ICs can be used to build a parallel 8-bit converter with a 100-Msample/sec (800-Mbps) word rate. These 4-bit quantizers make it feasible to build relatively small, medium-accuracy parallel type converters for use in aerospace systems.

A number of companies are employing LSI circuit technology to produce high-performance and/or low-cost monolithic converters. Several companies are producing single-chip 8-bit successive approximation converters that sell for less than $20.00 in small quantities. TRW has recently announced a family of ultra-high-speed analog-to-digital converters that employ LSI technology. This family includes an 8-bit converter with a 50-nsec conversion time and a 10-bit converter with a 100-nsec conversion time.

TRW is also developing a 4-channel, 5-bit, 400-Msample/sec data system for the Air Force Avionics Laboratory. The system contains a parallel type analog-to-digital converter packaged as a single hybrid module that uses four 3-bit quantizer chips and an emitter-follower differential logic family that has gate delays of 250 to 350 psec per gate.

4.1.1.2 Intra-Spacecraft Communications -

4.1.1.2.1 Serial Time Division Multiplexed Data Buses - During the past 10 to 15 years, serial digital time division multiplexed data bus communications techniques have played an important role in space data handling system designs. Most general-purpose data systems utilize a serial biphase encoded transformer coupled data bus as the communications link between the various components of the systems. These systems generally operate the bus at data rates of 1 or 2 MHz. Very large data systems such as used on the Shuttle orbiter employ multiple data buses. Multiple buses can be used either for redundant communications or to increase the data rate of the system.

Serial data bus designs were first utilized during the 1960's on various missile systems such as the Titan and Agena in an effort to reduce the weight of the instrumentation system. In 1969, the A2K Committee of the Society of Automotive Engineers was formed to investigate and develop standards for aircraft multiplexing. The work of the A2K Committee eventually evolved into MIL-STD-1553A, which was adopted by the Department of Defense in 1975. The initial application of multiplexing to commercial aircraft was the passenger entertainment system for the Boeing 747. The F-15, which was first test flown in 1972, was the first operational fighter aircraft to employ an operational 1-MHz data bus as the primary avionics system communications medium. The F-15 flight test system also utilized a serial data bus. The Space Shuttle has several types of serial data buses. The Shuttle Orbitor has a 1-MHz bus similar to the MIL-STD-1553A data bus. The Spacelab and External Tanks also use similar data bus techniques.

Most data bus designs employ transformer coupling with balanced twisted shielded pair cable to allow the system to operate in extremely high common-mode noise environments. Another technique sometimes used to achieve high isolation and common-mode rejection is the use of optical-isolators. Because of the high current requirements and temperature instabilities, optically coupled devices are generally not used in the larger systems that require multiple (up to 32) modems on a single data bus.

4.1.1.2.2 Fiber Optic Data Buses - The fiber optic data bus has been gaining widespread acceptance since about 1970. Fiber optics offer the advantages of small size, light weight, extremely high data handling capacity, high security from jamming, high immunity to EMI/EMP, crosstalk immunity, potential low cost, and lower transmission loss per length of cable. Both the Navy and the Air Force have sponsored several fiber optic demonstration programs and are currently engaged in developing components and modules as well as systems. The Navy is working on devices for second-generation 10-Mbit/sec systems. The Air Force is concentrating on the development of low-cost electrical-to-optical and optical-to-electrical interface circuits. The Navy program is being managed by the Naval Electronics Laboratory and the other program by the Air Force Avionics Laboratory. The outcome of these programs should significantly affect the role of fiber optics for future data bus applications in general-purpose data handling systems.

Some fiber optic demonstration systems already tested by the Navy include: a secure telephone system that was installed on the USS Littlerock in 1973, a fiber optic sonar link (FOSL-1) that consisted of 52 parallel channels to provide analog signal transmissions, and a 10-Mbit airborne optical data multiplexer designed to MIL-E-5400P standards for the Naval Ocean Center's Airborne Light Optical Fiber Program (ALOFT) and test flown on the A-7 aircraft for 150 hr. As a result of the ALOFT demonstration program success, the Navy is considering a point-to-point fiber optic system for the Advanced Harrier. In addition to these military demonstration programs, numerous fiber optic data links have been implemented for ground communciation applications. These are discussed in Section 9.

A fiber optic data link consists basically of a light-emitting source for optical fiber transmission media and an optical receiver. Each of these areas is discussed below.

There are three basic types of optical fibers:

1. Single-mode step index - A core small enough (1 to 10 μm) to contain only the lowest mode is bound by an outer cladding of higher refractive index, resulting in an abrupt increase in index at the interface.

2. Multimode step index - This fiber resembles the single-mode step-index fiber but has a much larger core (30 to 70 μm).

3. Graded index - The refractive index varies directly with radius in the graded-index glass fiber, which consequently exhibits low loss and low dispersion.

Figure 4.1.1.2-1 shows the relative conceptual differences between these basic fiber types.

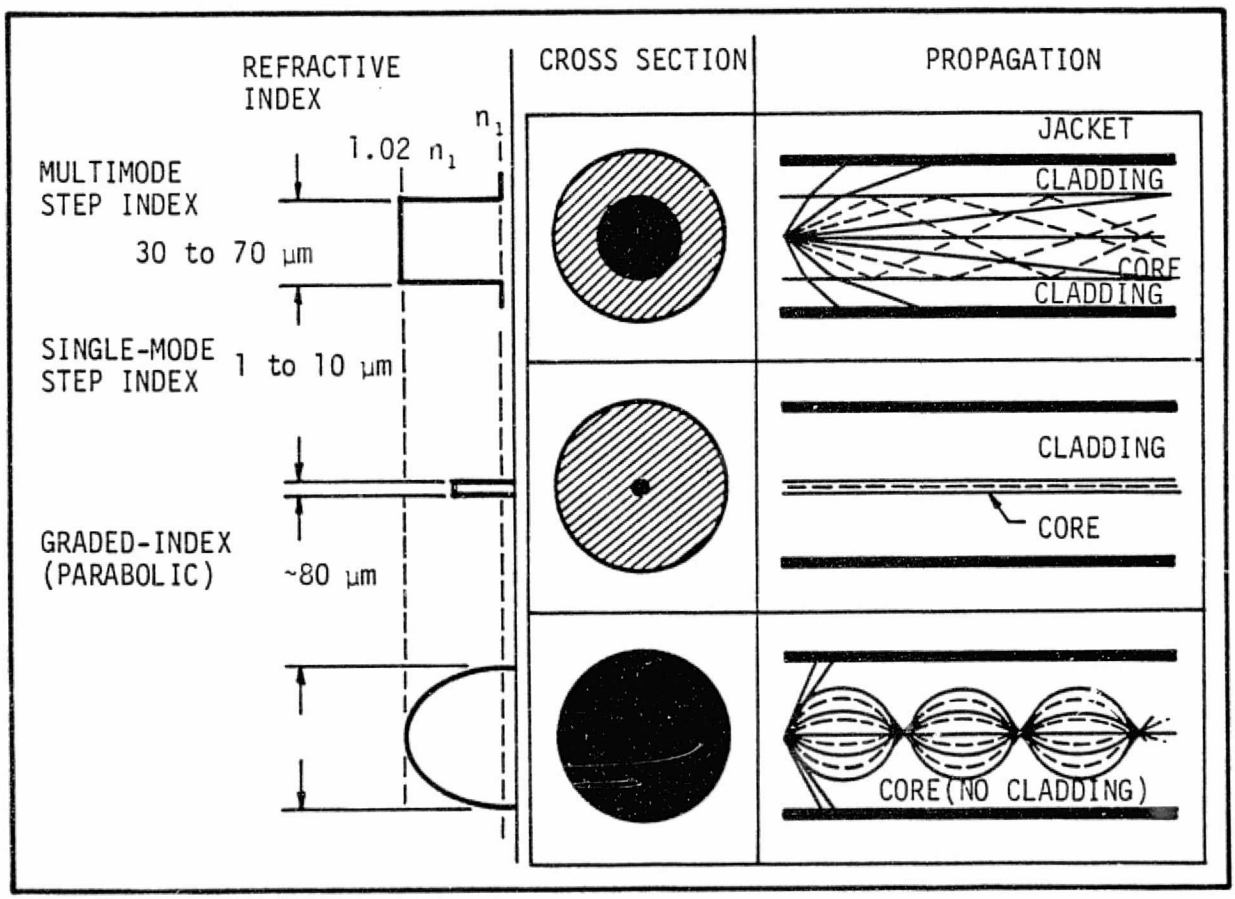


FIGURE 4.1.1.2-1. PRINCIPAL FIBER TYPES AND THEIR CHARACTERISTICS

For any optical communications application, fiber attenuation, fiber dispersion, and cost generally determine the type of fiber to be used. Attenuation loss in optical fibers is due to material absorption, material scattering, waveguide scattering, and radiation loss. For any fiber material there will be spectral regions of low and high loss. Figure 4.1.1.2-2 gives a typical attenuation curve for graded-index glass fibers as a function of wavelength. The graded-index fibers have the lowest attenuation loss but are more costly than the step-index types. Attenuation loss in production quality graded-index fibers has been as low as 5 dB/km for glass fibers whose spectral region of lowest loss is between 0.85 and 0.90 μm. The plastic multimode fibers are the lowest cost but have higher loss and dispersion than the other types.

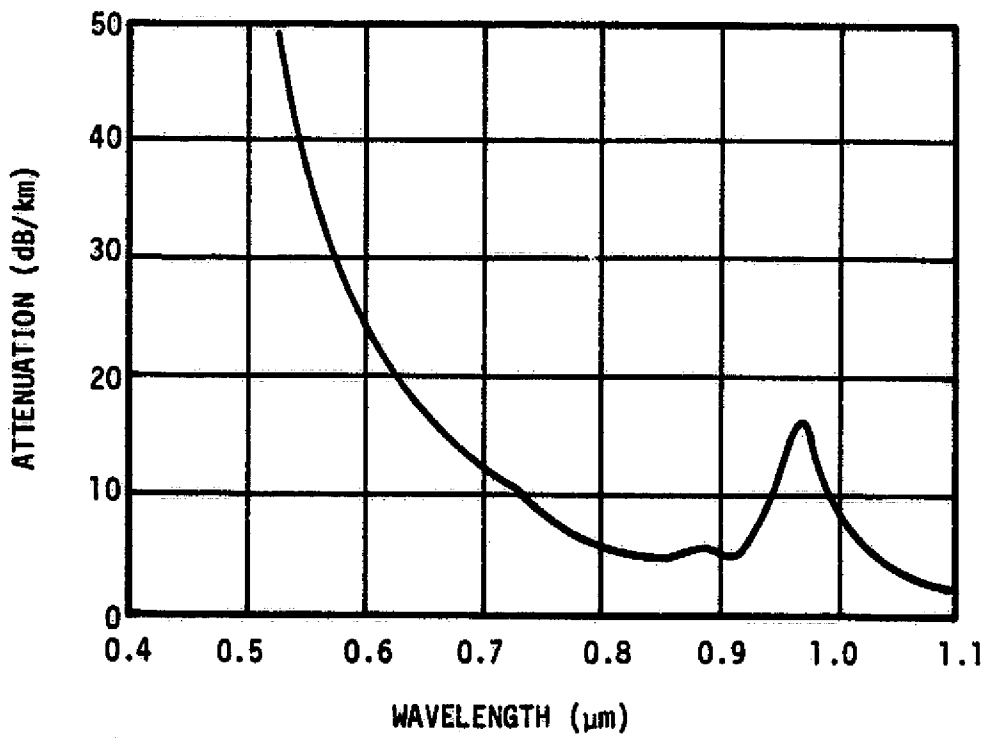


FIGURE 4.1.1.2-2. TYPICAL SPECTRAL ATTENUATION-GRADED INDEX FIBER

Another important fiber characteristic is optical dispersion, which is a measure of the spreading of optical pulses per unit length due either to the material or multimode group delay. The best value of dispersion for production quality graded-index fibers is 5 nsec/km. Experimental fibers have exhibited dispersions as low as 0.2 nsec/km. It should be noted that for relatively short distances, such as encountered onboard spacecraft, neither attenuation nor dispersion should limit system data rates.

Electrical data bus systems provide a capability for connecting multiple terminals to a single electrical bus. This is possible because the data bus receivers reflect a high impedance (~10 K) to the bus and the data bus transmitters function as active elements and are only "connected" to the bus when the data terminal is transmitting.

The current "first-generation" optical data bus systems employ either an optical tee at each data terminal or use a star coupler. The tee configuration minimizes cabling but can provide only a limited number of terminals without signal amplification. The star coupler configuration employs a single optical mixer or "star" to which all of the data terminals are connected. A comparison of worst-case loss as a function of the number of terminals for tee and star systems is shown in Figure 4.1.1.2-3.

Source characteristics for fiber optic communication systems are adequate output power, ease of modulation, long lifetime, high efficiency, low cost, and fiber compatibility. Three sources with these general characteristics are:

- Semiconductor light emitting diodes (LED)
- Semiconductor injector lasers
- Solid-state lasers.

In LED structures, the optical output is emitted over a solid angle of 2π sr. Therefore, even if the device has high output power, its output radiance, or output power emitted into a unit solid angle per unit emission area, can be low, resulting in high coupling losses. Emphasis is placed on high radiance output in LEDs to overcome these losses.

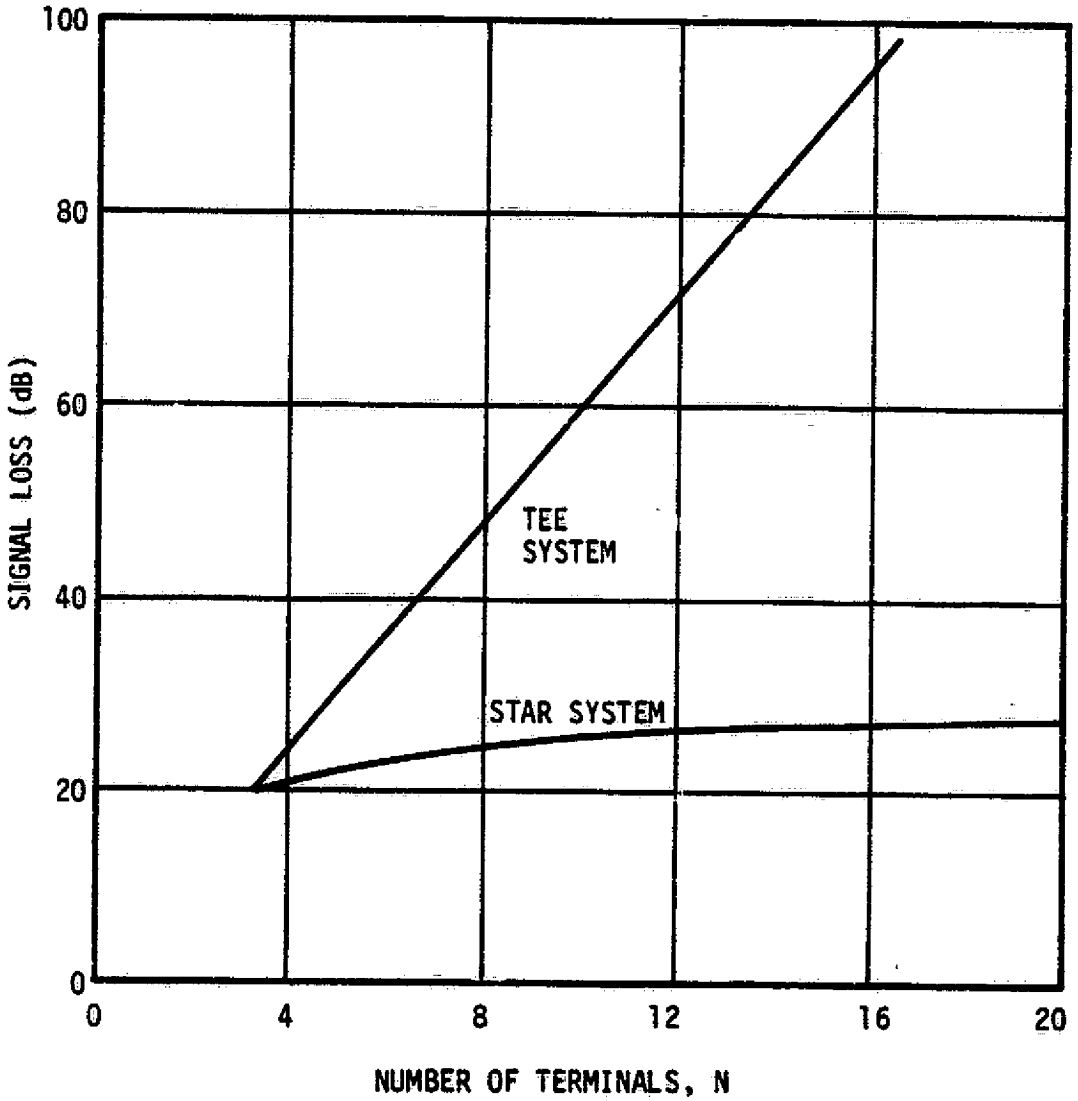


FIGURE 4.1.1.2-3. COMPARISON OF COUPLING LOSSES FOR STAR AND TEE CONFIGURATIONS

High current density, thermal shock, and mechanical stress can all limit the lifetime of an LED. Current LEDs have been run up to 3 years, or 25,000 hr, in test environments. The results of accelerated life testing indicate that lifetimes in excess of 100,000 hr are possible.

The output spectral width is inversely proportional to the material dispersion of the fiber. LEDs have a spectral output width of 40 nm in the best case, approximately an order of magnitude wider than injection lasers.

Optical energy generated by injection lasers is essentially limited to a 30-deg cone angle so that a much larger fraction of energy may be coupled into the multimode fibers than is possible with LEDs; hence coupling losses are lower for injection lasers than LEDs.

Injection lasers can be directly junction-modulated by varying the injection current, thus providing wide modulation bandwidths and faster data rates. Injection lasers have an output spectral bandwidth that is a factor of ten better than LEDs: 4 nm.

The most highly developed semiconductor material for injection laser use is aluminum-doped gallium arsenide, which operates in the spectral region of low loss for glass fibers: 0.80 to 0.85 μm. Other semiconductor materials are being developed to operate further into the infrared. Research is also being performed in developing low-loss fiber materials to operate in the far infrared.

The Nd:YAG laser is a solid-state laser that requires optical excitation in pump bands located at $\lambda = 0.75$ and 0.81 μm. It has been considered for fiber optical applications because its laser emission at $\lambda = 1.06$ μm is in the low-loss range of glass fiber attenuation. Its output spectral width is less than 0.1 nm, resulting in very little material dispersion (see Table 4.1.1.2-1). However, the device must employ external modulation because the upper energy levels have a long fluorescence lifetime of about 200 μsec. Consequently, direct laser modulation is limited to bandwidths below megahertz rates, thus restricting the data handling rate. Miniature Nd:YAG lasers that are compatible with optical fibers have been fabricated using LED end-pumping schemes.

TABLE 4.1.1.2-1. STATE-OF-THE-ART FIBER OPTIC COMMUNICATIONS COMPONENTS

COMPONENT	PROPERTY	PRODUCTION	RESEARCH
Fiber Optics	Pulse Spread	5 nsec/km	0.2 nsec/km
	Bandwidth	400 MHz	1 GHz
	Attenuation	5 dB/km	1 dB/km
Detectors	NEP	10^{-12} to 10^{-14} W	10^{-16} W
	Rise Time	10^{-9} sec	$<10^{-12}$ sec
	Data Handling	700 Mbit/sec	1 Gbit/sec
Sources			
LEDs	Spectral Width	40 nm	40 nm
	MTBF	10^{4} hr	10^{5} hr
Injection Lasers	Spectral Width	4 nm	4 nm
	MTBF	10^{4} hr	10^{5} hr
Nd: YAG	Spectral Width	<0.1 nm	<0.1 nm
	MTBF	Not Available	Not Available

A typical optical receiver consists of a photodetector, a front-end amplifier, a filter, and demodulation circuitry. For most fiber applications, the detector is usually a p-i-n or avalanche photodiode. The photodiode response time is the rms sum of the charge collection time in the depletion region of the semiconductor and the RC time constant produced by the load resistance and the sum of the junction capacitance and stray capacitance. The response time, or rise time, describes the speed with which a detector responds to a change in the incident flux and hence limits the data rate. The stray capacitance (assuming low load resistance) is the limiting factor in the response time. The capacitance can be reduced, however, by reducing the active surface area. This causes a shorter rise time but in so doing decreases the overall coupling ability of the photodiode. In this sense, the response time can be a deceptive figure, and it is necessary to understand which values

of load resistance and active surface area the manufacturer is implying in his response time quotations. Current values for rise time are less than a nanosecond. Some present-day photodetectors have demonstrated a data handling rate as high as 700 Mbits/sec, but with low signal to noise ratio.

The noise equivalent power (NEP), or the minimum incident power required to generate a photocurrent equal to the total photodiode noise current, is of great importance because it defines the minimum detectable power of the photodiode. Like the response time, the NEP can be an ambiguous figure if the test conditions are not specified. A manufacturer's NEP figure implies knowledge of source power, chopping frequency of the radiation, and noise bandwidth, as well as the previously mentioned values of resistance and capacitance. The NEP for current photodiodes is 10^{-12} to 10^{-14} W.

Characteristics of state-of-the-art fiber optic components are presented in Table 4.1.1.2-1.

4.1.1.3 System Control - The data system control function can either be integrated with the signal conversion hardware, as in small PCM systems, or it can be a separate hardware element, as in the larger data systems. Data system control functions include the following:

- Data selection (sampling sequence generation)
- Internal communications control
- Data interleaving and output formatting
- Data buffering.

4.1.1.3.1 Data Selection and Internal Communications Control - State-of-the-art data systems provide control of the input data sampling sequence by either hardwired logic or program control. Data acquisition systems with relatively simple data formats use either hardwired logic or a combination of hardwired logic coupled with some type of externally wired program plug. The larger, more complex data systems generally use some type of centralized controller that stores the data sampling sequence on a programmable memory.

In the past, most spaceborne data systems have utilized a system control unit that is separate from the onboard computer to acquire housekeeping and telemetry data. Frequently, the guidance, navigation, and control system have used a separate data system dedicated to these functions. State-of-the-art spacecraft data system designs integrate the data acquisition and control distribution functions into a single system. In some of these systems, the data acquisition process is controlled by a dedicated processor or controller. In other configurations, the controller is implemented as a computer I/O channel that shares memory with the central computer by cycle stealing to generate the data sampling sequence and control the flow of commands and data between the controller and the various system data acquisition terminals. Both types of controllers are used in the various Space Shuttle data systems.

In systems that have a separate system control unit, interfaces are usually provided between the system control unit and the onboard processors, onboard data storage units, and the communications downlink. In general, the technology used to implement system control units is the same as used in spaceborne processors with the single exception that

greater emphasis is placed on low-power operation for the controllers. The computer technology discussed in Section 2 also applies to data system controllers.

4.1.1.3.2 Data Interleaving and Output Formatting - Asynchronous multiplexing, data interleaving, and data buffering has within the last few years become an important aspect of spaceborne data system design. The Orbiter Payload Data Interleaver and the Spacelab High Rate Multiplexer are examples of these types of hardware. The High Rate Multiplexer is the most advanced general-purpose high-rate digital multiplexer development reported using commercially available components. The High Rate Multiplexer is a 50-Mbit/sec general-purpose asynchronous multiplexer being developed for NASA by Martin Marietta. This system has 16 multiplexer inputs, each of which will accept up to 16 Mbits/sec. However, the maximum combined system input data rate is 48 Mbits/sec. The system is coded in NRZ-L format with separate clock lines to each input. The system is designed using technology capable of being space-qualified. The High Rate Multiplexer does not buffer blocks of data for formatting purposes. Data are output and multiplexed a word at a time in a manner similar to conventional data systems in the past. Fill bits are output when data are not available from the asynchronous inputs.

One of the major problems in high-rate digital data handling is power consumption. One technique for reducing power in aerospace data systems is to operate parallel subsystems constructed from components with lower power-delay products and to combine these subsystems outputs using a minimum of high-power logic to produce a high-speed system output. Another way to reduce power is to attack the basic physics of the power consumption problem, which is the amount of energy required to charge device capacitances. The power can be reduced by reducing either voltage swings or device capacitances. Harris used custom-designed, large-scale ECL logic in MOMS to reduce both the internal logic voltage swings and the integrate capacitances. Typical power-delay products achieved were an order of magnitude less than MECL III for 200-MHz operation.

Several new logic technologies with extremely low-power delay products are discussed in Sections 1 and 7. These include enhancement and depletion mode GaAs MESFET, low-power silicon MESFET, static induction transistor logic, and elevated electrode logic.

Another system utilizing custom-designed ICs is a high-rate digital multiplexer developed by TRW Systems for the Air Force Avionics Laboratory. The multiplexer has six input channels. Each channel multiplexes 8 bits of parallel data at a 20-MHz word rate. The system adds the synchronization code and converts the data to a 1-GHz serial data stream. The logic technology employed in the system is TRW's emitter-follower differential logic with oxide-lined transistors and junction isolation. Typical gate delays are 250 to 350 psec.

Subsection 9.1.1.2 reports on a 274-Mbit/sec multiplexer (AT&T's D34 multiplexer) that began operation on a test basis between New York City and Newark, New Jersey, in 1975. The D34 system accepts three 44.736-Mbit/sec pulse streams into each of two 137-Mbit/sec multiplexers and interleaves the two 137-Mbit/sec bit streams at the last instant to minimize the requirement for high-speed (≈1-nsec) logic. This system is not restricted in power consumption, weight, size, etc., as in a space-borne system. However, as reported in Section 9, the same types of ICs were used in this system as in the system discussed below, which is being built for space data-handling purposes.

Another high-rate data system that uses commercially available components is a 200-Mbit/sec simulator built by Martin Marietta in 1974 (Ref. 4-2). Work is currently being performed on a 500-Mbit/sec system that uses faster components than were available when the 200-Mbit/sec system was built. The 200-Mbit/sec system was designed with Motorola MECL 10,000 and MECL III components and was successfully operated at 230 Mbits/sec under varying signal conditions. MECL III components were used where speed was a primary consideration and either MECL 10,000, TTL, or CMOS components were used where possible to minimize power consumption. The engineer responsible for the program was contacted to determine the power consumption and the feasibility of this high-data-rate system for space applications. The actual power consumption was

not established, but the opinion was that the consumption of power was too high for space applications.

At the time 200-Mbit/sec system was built, the state of the art in available logic components permitted a toggle rate to approximately 350 MHz using the Motorola MC 1670 type D flip-flop from the MECL III series. Presently, Fairchild semiconductors offer flip-flop components capable of toggle rates to approximately 720 MHz. Thus improvements in speed by a factor or two over previous systems are reasonable to expect.

4.1.2 Trends in Digital Data Handling Systems

The trends in general-purpose digital data systems are toward increased capability and flexibility, smaller size, and lower power consumption. Flexibility will continue to be enhanced in future systems as microprocessors and memories with greater capacities are incorporated into the designs. Bubble memories (Subsection 3.2) will probably replace plated wire memories and EPROMs for nonvolatile program storage in future system format controller designs.

Future general-purpose data systems will utilize microprocessors within the remote terminals to perform functions such as data queueing, data compression, and buffer storage. These intelligent remote terminals will significantly reduce the amount of traffic on the data bus by eliminating much of the nonessential data transmitted on the bus.

Unique/dedicated data system trends are toward higher data rates and the application of onboard processing to control selective sampling, control sensors, perform data compression, and provide limited onboard real-time data analysis. Although there is a current trend in unique/dedicated data systems toward higher data rates with the Landsat-D Thematic Mapper data rate almost six times the rate of the Landsat-C Multispectral Scanner, the data system needs for the immediate future do not provide a significant technological challenge. The technology currently exists to build data systems having 8-bit resolution with output data rates up to 1 Gbit/sec. Current developments in LSI analog-to-digital converter technology, GaAs metal-semiconductor field-effect transistor logic technology, and short-channel NMOS technology for analog switches should make possible the construction of even faster data systems by 1985.

Although the technology is available to build much higher rate data systems than currently exist, it is unlikely that these types of systems will be utilized by NASA in the immediate future since ultra-high-rate space-to-ground data links do not currently exist and none are planned that will accommodate in excess of 600 Mbits/sec. It may therefore become necessary beyond Landsat-D for onboard data systems to

include a significant amount of data processing to selectively sample and/or perform data compression.

Many of the current developments in data systems circuit technology are being driven by radar processing and electronic countermeasures applications. This includes both ultra-high-speed analog-to-digital conversion and onboard real-time processing. Several companies, including Harris, IBM, Motorola, TRW, and others, are developing high-rate data systems for radar processing. JPL has a 5-year program to develop a synthetic aperture radar (SAR) processor. Possible applications for the SAR processor include SEASAT-C, Spacelab, and Venus Orbiter.

4.1.2.1 Signal Conversion - Many authorities predict the continued demise of sensors and subsystems that output analog signals to the data system. Although this trend has already eliminated most of the analog type outputs from the larger spacecraft subsystems such as inertial platforms and experiments, there are a number of types of onboard sensors (e.g., temperature and pressure) that most likely will continue to provide analog inputs to the data system and require some type of signal conditioning.

Increases in the speed and complexity of integrated circuits, decreases in power consumption, and decreases in cost for complex chips can be expected to continue for the foreseeable future (Subsection 7.2.1), and the effects on space data handling systems will continue to be felt. Another trend that will impact signal conversion technology is the increased use of MOS/CMOS linear devices in future high-speed multiplexer and analog-to-digital converter applications as the performance and cost of these devices challenge their bipolar counterparts. It has been known for many years that majority carrier devices (such as NMOS) should be faster than minority carrier devices (such as bipolar transistors). However, only recently have short-channel MOS devices that were faster than bipolar devices actually been fabricated. The integration of short-channel MOS technology into analog multiplexing and analog-to-digital converter subsystem designs will provide significant improvements in the speed of these devices.

Several companies, including National Semiconductor and Texas Instruments, have recently introduced BI-FET operational amplifiers that combine the advantages of a high-impedance FET input with a bipolar output on a single chip. National is also applying BI-FET technology to other signal conversion applications including analog swtiches, comparators, and sample-and-hold circuits.

The trends in monolithic analog-to-digital converters are toward lower cost, higher resolution and accuracy, and faster conversion rates. Figure 4.1.2.1-1 presents the cost trend for 8- and 12-bit monolithic and hybrid successive approximation analog-to-digital converters projected through 1985.

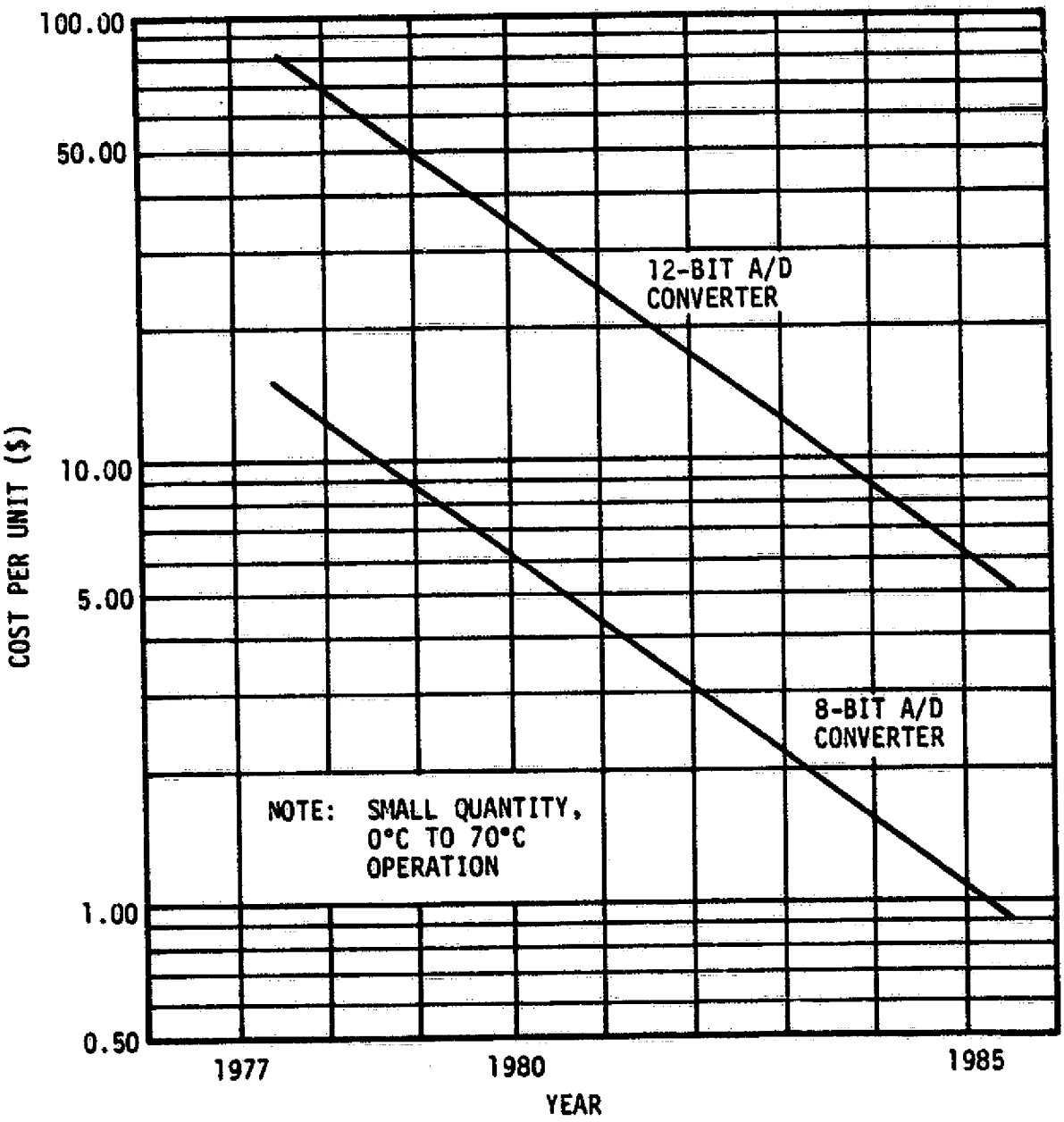


FIGURE 4.1.2.1-1. COST PROJECTION FOR MEDIUM-SPEED, COMMERCIAL-QUALITY, MONOLITHIC/HYBRID ANALOG-TO-DIGITAL CONVERTERS

4.1.2.2 Intra-Spacecraft Communications - Flexible data system designs utilize standard intra-spacecraft communications interfaces to achieve modularity. Although most state-of-the-art data systems utilize some type of standard interface bus between the various subsystem components, there are indications that simple standardization will not be sufficient to provide future system growth capacity required to accommodate rapidly evolving sensor and processor technology. It is anticipated that future data bus designs employing fiber optic technology will be structured to provide "growth-oriented" modularity such that the system communications throughput rates can be increased as necessary without redesigning the subsystem interfaces on the system data bus.

Figure 4.1.2.2-1 projects the trend in serial data bus data rates through 1985. It is anticipated that the current MIL-STD-1553A twisted shielded pair 1-Mbit/sec data bus will be supplemented by a MIL-STD-1553A format-compatible for 10-Mbit/sec fiber optic bus by 1980 and that the data rate will be extended to 100 Mbits/sec by 1985. Systems requiring data rates in excess of 100-Mbits/sec will most likely employ parallel buses.

The availability of optical fibers with low attenuation (less than 20 dB/km) has stimulated significant commercial interest in the technology for ground point-to-point communications. Figure 4.1.2.2-2 shows the trend in attenuation characteristics of fiber optic materials through 1985. As seen in Figure 4.1.2.2-2, optical fiber attenuations are currently lower than the best coaxial cable at 10 MHz. As a result of the widespread commercial interest in fiber optics, it is anticipated that this technology will continue to evolve at a rapid pace, resulting in the development of commercially available low-cost, high-stability optical fibers with attenuations of 1 dB or less per km by 1985. Optical fibers having attenuations of less than 0.5 dB/km have already been demonstrated in the laboratory. As previously illustrated in Figure 4.1.1.2-2, graded-index glass fibers have the lowest attenuation in the infrared region of the spectrum. Improvements in manufacturing techniques

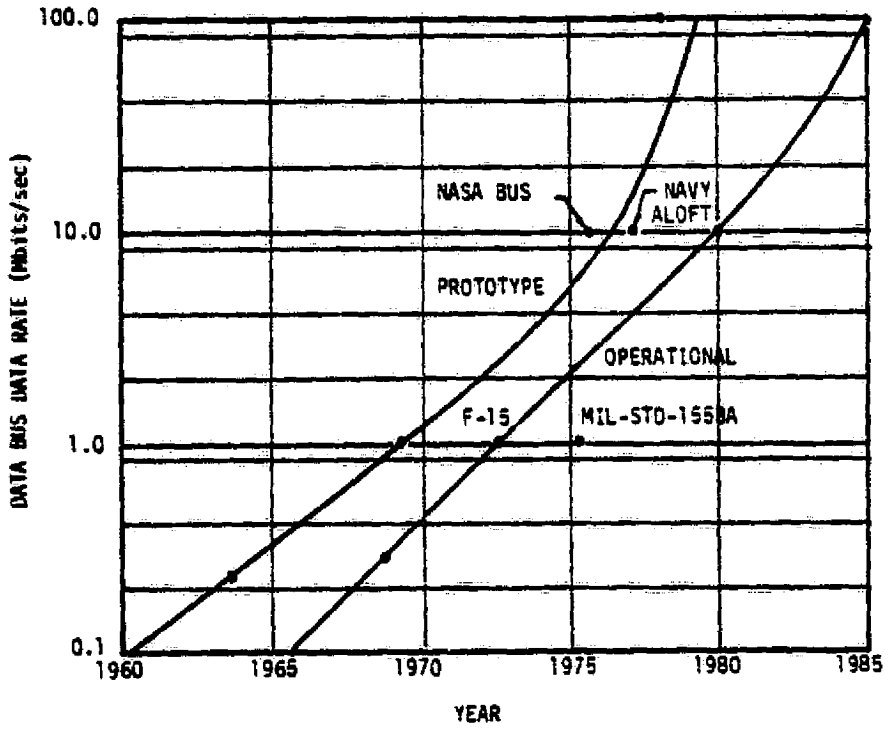


FIGURE 4.1.2.2-1. TRENDS IN SERIAL DATA BUS DATA RATE

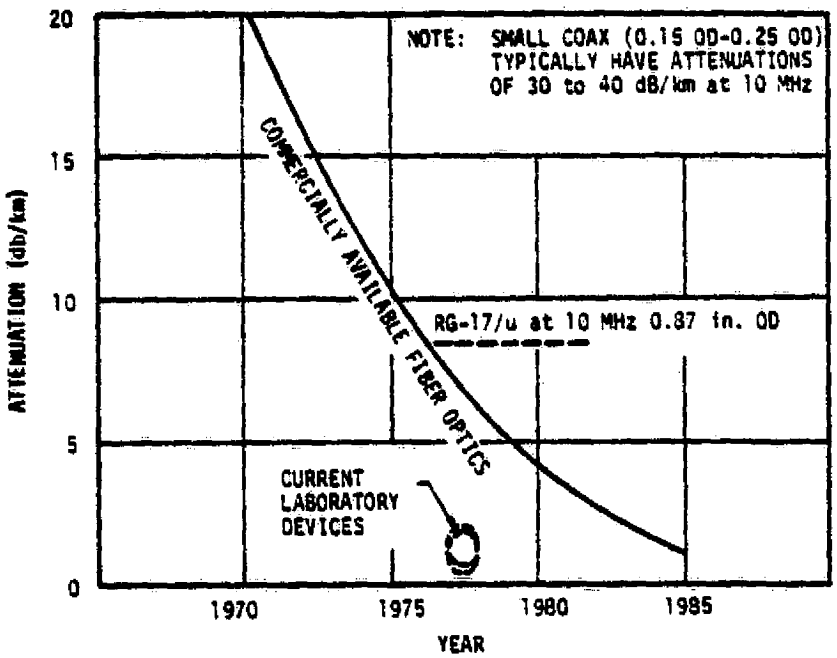


FIGURE 4.1.2.2-2. TRENDS IN OPTICAL FIBER ATTENUATION

should produce higher purity fibers. Graded-index glass fibers have been produced in the laboratory with a dispersion of 0.2 nsec/km. This will most likely be the state of the art in fiber dispersion for 1985.

Research is being performed in developing fibers of different materials to operate in the far infrared (6 to 35 μm) with improved characteristics. There is little available source technology in this spectral region, as compared with visible and near infrared regions previously discussed.

Research is currently being performed to reduce the solid angle of output radiance by altering the semiconductor configuration, but major improvements are not expected in LEDs by 1985 because of the emphasis being placed on injection lasers as sources for optical communication systems.

Semiconductor materials are being developed to operate further into the infrared. Technology for fibers and emitters in the far infrared region is seriously lagging the near-infrared advances, and no major breakthroughs are anticipated by the mid-1980s. Injection lasers operating in a spectral band commensurate with spectral regions of lowest loss for glass fibers will be the most likely optical communications source in 1985.

If the modulation problem of the Nd:YAG laser can be solved, the Nd:YAG laser will likely become the preferred source for fiber optic communication systems. However, in view of the complexity of the laser/modulator combination compared with the success of semiconductor devices, it appears that the Nd:YAG laser will not be realized as a major technology in optical communications by 1985.

Photodetectors with data handling rates in excess of 10^9 bits/sec are expected by 1985, but resulting low signal-to-noise ratios and system design standards in effect at that time may force designers to use systems with lower data handling rates.

4.1.2.3 System Control - Two current trends in data system control hardware are distributed control and asynchronous data handling. Distributing the control function to the various subsystem elements reduces the data bus traffic in general-purpose data systems by eliminating the need to transmit control or sampling commands to the remote data terminals. NASA and JPL are currently planning the development of a Remote Terminal Unit (RTU) that will initially contain an 8-bit microprocessor with 1K words of programmable read-only memory (program storage) and 2K words of random-access memory (buffer memory). A simplified block diagram of the RTU is shown in Figure 4.1.2.3-1. It can be assumed that as more complex microprocessor chips and memory chips become available, more sensor processing will be incorporated into future systems designs.

NASA has instituted a Low Cost Systems Office for the purpose of promoting hardware standardization throughout NASA. Hardware such as the RTU will become a "standard" component to be used throughout the various programs of the various NASA centers. This trend toward hardware standardization should provide significant cost savings for future NASA data systems.

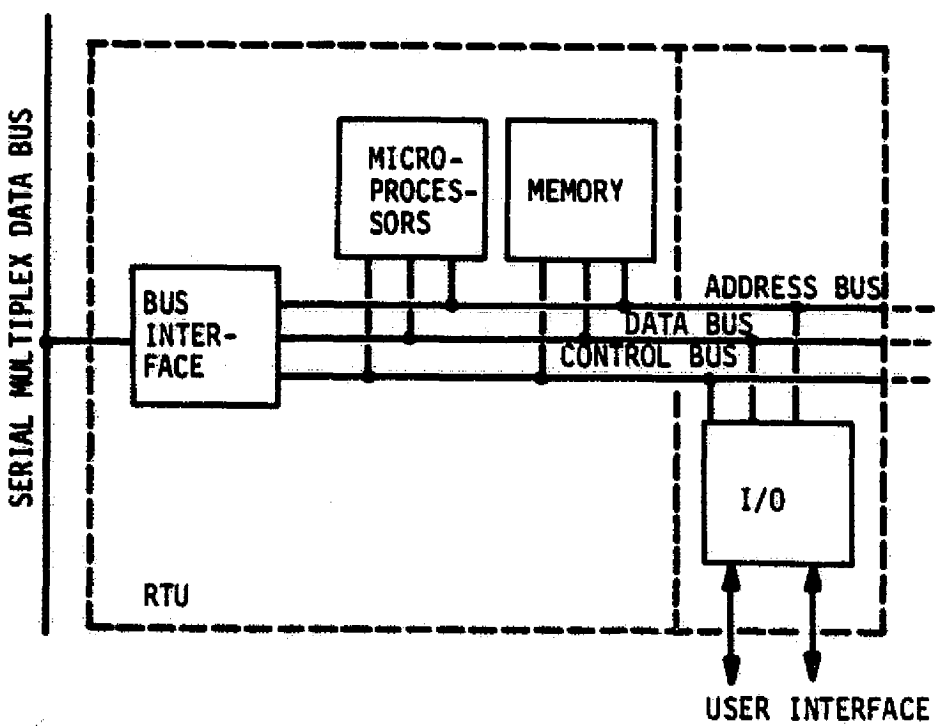


FIGURE 4.1.2.3-1. REMOTE TERMINAL UNIT BLOCK DIAGRAM

GSFC has initiated another program to develop and flight qualify standardized equipment for the performance of onboard telemetry and command functions. The basic elements of the Standard Telemetry and Command Components (STACC) are illustrated in Figure 4.1.2.3-2. The STACC system has a multiplexed data bus for remote distribution of commands and remote multiplexing of telemetry data. Other characteristics of the STACC equipment are:

- Onboard computer interface to provide communications between the computer and all spacecraft subsystems and experiments
- Two command-selectable telemetry (TLM) formats controlled by read-only memories, plus variable formatting through use of an onboard computer
- Fully redundant multiplex data bus
- All components can be used in redundant configuration.

STACC equipment is currently under development and should be available during 1978. The first scheduled user of the STACC equipment will be the Solar Maximum Mission, which is scheduled for launch in the second half of 1979.

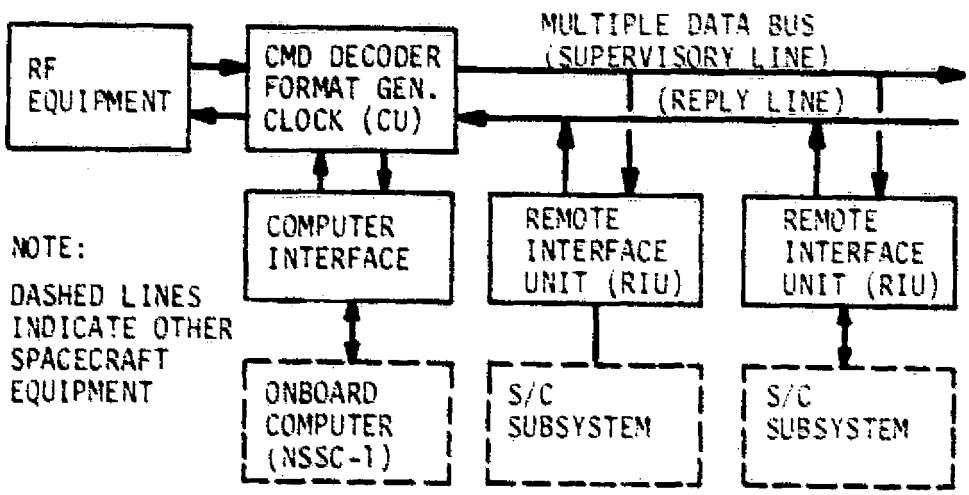


FIGURE 4.1.2.3-2. STANDARD TELEMETRY AND COMMAND COMPONENTS

A data compression technique that will likely be implemented in next-generation spaceborne systems is "data-set selection". This is a technique for screening and selecting only useful data sets for onboard processing or transmission to the ground. For example, in the case of Landsat imagery, data containing excessive cloud cover might be completely rejected.

The application of distributed processing to spaceborne data system designs will result in the increased use of asynchronous data handling techniques employing packet-switched data. These priority-controlled data handling systems will use store and forward techniques to provide variable bandwidth capabilities for more efficient management of the available data link capabilities.

4.1.3 Projected Developments in Digital Data Handling Systems

4.1.3.1 General-Purpose Systems - Projected developments in general-purpose digital data systems are toward flexibilities and capabilities heretofore unavailable. The most significant development will be the ability by 1985 to package a powerful computer with a large memory (on the order of 10 Mbits) on a single small PC board, thus making it possible to perform complex processing functions within each remote data terminal.

Another projected development is an increased requirement for asynchronous multiplexing and associated data buffering. Low-cost bubble memories and/or CCD memories will be used to provide buffer storage for asynchronous data interleavers and temporary data buffers. As discussed in Section 3, the availability of relatively small bubble memory devices with in excess of 10^{10} bits will make data-set selection and other limited analysis techniques feasible for next-generation data systems. Table 4.1.3-1 compares some typical characteristics of current general-purpose data systems and general-purpose data systems projected for general use by 1985.

As seen in Table 4.1.3-1, the major difference between the future and current data systems is the transformation of the intelligence or processing to near the source or sensor. Another major difference in projected capabilities of general-purpose data systems is much higher intra-spacecraft communications data rates as a result of fiber optic technology.

4.1.3.2 High-Rate Systems - The data system for the Thematic Mapper is probably the most advanced 8-bit data-handling system currently under development for advanced NASA programs. Currently, NASA has no exploratory data system development work for data systems beyond Landsat-D. The Air Force system previously discussed in Subsection 4.1.1 represents the highest-data-rate (2-Gbps) system currently under development employing a single analog-to-digital converter. High-speed GaAs MESFET logic currently achieves 100-psec propagation delays in laboratory tests, and

TABLE 4.1.3-1. COMPARISON OF TYPICAL CURRENT AND FUTURE GENERAL-PURPOSE DIGITAL DATA SYSTEMS

CHARACTERISTIC	CURRENT (1977)	1985
System Type	Modular, stored program, party line	Modular, stored program, party line
Data Output Rate	500 kbits/sec	5 Mbits/sec
Party Line Construction	Twisted shielded pair	Fiber optics
Party Line Rate	1 Mbits/sec	100 Mbits/sec
Party Line Operation	Command-response	Asynchronous demand
Remote Unit Capacity	32	32
Central Unit Processing Capability	Sample format, gain offset	Remote unit parity, data queueing, security encoding, correlative analysis, etc.
Remote Unit Programming Capability	Remote unit address	Channel (sample rate), offset, data distribution, data compression, self-test, limited analysis, etc.

frequency dividers have been operated with actual counting rates of 4 GHz. Projections of current device technology trends to 1985 result in the high-rate unique/dedicated data system characteristics presented in Table 4.1.3-2.

TABLE 4.1.3-2. HIGH-RATE DATA SYSTEM TECHNOLOGY PROJECTIONS FOR 1985

Analog Multiplexer Switch	Short-channel NMOS
Analog Switching Speed	<400 psec
Analog-to-Digital Converter	All parallel for 4 to 8 bits, series-parallel for 8 to 12 bits
Conversion Speed	2 nsec for parallel converter, 10 nsec for series-parallel converter
Serial Output Data Rate	To 5 Gbits/sec
Ultra-High-Speed Logic	GaAs MESFET, elevated electrode logic

The response from authorities and the results of in-house analysis indicate that progress in commercially available high-rate data handling systems will continue, but possibly at a somewhat slower rate than has occurred over the past 5 years. The lack of more rapid progress over the next decade is based on a lack of emphasis in the area more than on a slowdown in the supporting technologies. What will happen in high-cost militarized systems cannot be established because of security reasons.

Some authorities predict that the highest-rate space-qualified general-purpose digital multiplexers that will be available as off-the-shelf components in the 1980-1985 timeframe will operate on the order of 100 to 125 Mbits/sec. Motorola has been producing shift registers in the MECL III series that operate in the 300-MHz range for the past

3 to 4 years. The device is based on circuitry with propagation delays on the order of 1 nsec. Subsection 7.2.1 reports the state of the art in ECL components as 0.5 nsec at present and projects improvements to approximately 0.2 nsec by 1980 and 0.1 nsec by 1985. As reported in Subsection 4.1.3.1, TRW is currently building custom logic with propagation delays as low as 0.25 nsec that operates at 1 GHz. The Air Force Avionics Laboratory (Ref. 4-4) projects that, by 1981, reasonably complex GaAs ICs will be available to operate at 5 GHz, with power levels of 5 to 10 mW per gate. In view of these projections, development of a 600-Mbit/sec data system in the immediate future appears feasible.

Reference 4-2 by John Goodwin emphasizes the importance of RF design techniques for these high bit rates. Others in the area also emphasize the same points. Thus, for data rates beyond 500 MHz, attention must be given to such other aspects of the system design as low-capacity printed-circuit-board materials such as Teflon and R. T. Duroid.

4.2 ANALOG DATA SYSTEMS

Analog data systems are discussed in terms of time-division-multiplexed (TDM) systems, frequency-division-multiplexed (FDM) systems, and hybrid systems that employ a combination of time and frequency multiplexing. Except for instrumenting development vehicles to obtain vibration and acoustic data, analog data handling systems have been replaced almost completely by digital data systems.

In general, other than minor packaging innovations, no major new developments have been made in analog data systems during the past 15 years. Since the state of the art in analog data systems is not changing as fast as it is for the digital data systems, this report emphasizes the digital system technology.

4.2.1 Frequency Division Multiplexing (FDM)

Most of the current FDM systems utilize IRIG-compatible FM subcarrier oscillators that are packaged as microminiature pluggable assemblies. The microminiature subcarrier oscillator package configuration that is most popular was first introduced by Vector (now Aydin Vector) in 1963. Several other companies, including Microcom and Omnitek, have developed somewhat smaller (submicrominiature) subcarrier oscillator designs. The Omnitek series 30 subcarrier oscillators are packaged in a standard 14-pin integrated circuit DIP configuration. These units are mounted by soldering to a PC board rather than plugging into a socket.

Future analog channel bandwidths will be increased significantly by the latest change in the IRIG Standards that are scheduled to be published later this year. The IRIG STD 106-77 expands the bandwidths of the baseband and increases the number and the bandwidth of the available channels. For example, the number of proportional channels will be increased from 21 to 25 and the highest center frequency changed from 165 kHz to 560 kHz.

4.2.2 Time Division Multiplexing (TDM)

Analog time-division multiplexing data systems used in the past have included pulse-amplitude modulation (PAM) and pulse-duration modulation (PDM) techniques. During the past 15 years, PAM systems have been replaced by PCM systems except for some small missile test applications that continue to use PAM. The IRIG STD 106-1975 deleted PDM because of a lack of usage. It is not anticipated that any spaceborne systems of the future will employ either PAM on PDM. Technology projections are therefore not given for these types of data systems. The analog multiplexers used in the digital data systems are basically identical to the multiplexers used in the PAM. However, since practically all PAM systems in use conform to these IRIG standards, it is anticipated that technology advances will only be incorporated in these types of systems to reduce hardware costs.

4.2.3 Hybrid Systems

Space data systems in the past have employed hybrid systems that included various combinations of frequency-division and time-division multiplexing in a single system. For example, one of the FM subcarrier channels on the Saturn V was modulated by a PAM time-division multiplexer.

It is not anticipated that hybrid PAM/FM analog systems will ever be employed in operational spaceborne systems; however, hybrid PCM/FM systems may be used in future booster/spacecraft test programs. The PCM/FM configuration would consist of an FDM system with one or more of the wideband FM channels modulated by a PCM wavetrain. The FM subcarrier channels would handle wideband (2-kHz) acoustic data, and the PCM system would handle low-frequency (50-Hz) vibration data. For this type of hybrid configuration, the technology discussions already presented in Subsections 4.1.1 and 4.2.1 apply.

REFERENCES - SECTION 4

4-1 W. T. Burton et al., "Multimegabit Operation Multiplexer System", Final Report, NASA Contract NAS5-21690, Harris, Electronics Systems Division, September 1973

4-2 J. E. Goodwin , "A 200 Megabit per Second Data Handling/Data Link Simulator", Proceedings of 1975 International Telemetering Conference, Volume XI, sponsored by International Telemetry Conference, pp. 102-111

4-3 C. Liechti et al., "Gallium Arsenide Spawns Speed", IEEE Spectrum, pp. 41-47, March 1977

4-4 "Military Integrated Circuits-Microcircuits Status and Trends-Technology Trends", Countermeasures, pp. 44-46, 59-60, May 1977

4-5 B. Abrahamsen et al., "A 560 Mbit/s Digital Transmission System Over Coaxial Cable", Proceedings of 1976 International Zurich Seminar on Digital Communications, sponsored by IEEE, pp. A4.1-A4.6

4-6 H. I. Maunsell et al., "The M13 and M34 Digital Multiplexers", Proceedings of 1976 International Conference on Communications, sponsored by IEEE, pp. 48-5, 48-9

4-7 B. Elson, "Fiber Optics' Aerospace Role Studied", Aviation Week and Space Technology, pp. 56-57, February 9, 1976

5. SPACE-TO-GROUND COMMUNICATIONS ELEMENTS

Data systems elements that make up the space-to-ground communication link are depicted in Figure 5-1. Level 2 hardware elements are discussed in terms of the principal components that establish their key characteristics, and Level 2 signal design techniques are presented in terms of relationships between the techniques and advances in technology. The primary emphasis throughout the discussions is on digital communication links since trends and major developments indicate an almost exclusive use of digital data in the future.

A major portion of the section is devoted to transmitting devices. This emphasis appears justified as a result of the trend toward greatly increased data rates, which results in a need for higher transmitter power and, in extremely high data rate cases, operation at a higher frequency. Advances in the technology of all communication devices are making communications at these higher data rates feasible; however, advances in signal-generating devices for transmitters appear to account for the greatest improvement of any single area.

The section addresses data relay satellites as one element of the space-to-ground link. The coverage of relay satellites in this section is restricted to wideband links between Earth-orbiting vehicles and the ground. Point-to-point transfer of ground telecommunications data via relay satellites is addressed in Section 9. In both cases, the information is presented in terms of the capabilities of planned satellites within the time periods of interest in lieu of discussing the technology elements that make up the satellites.

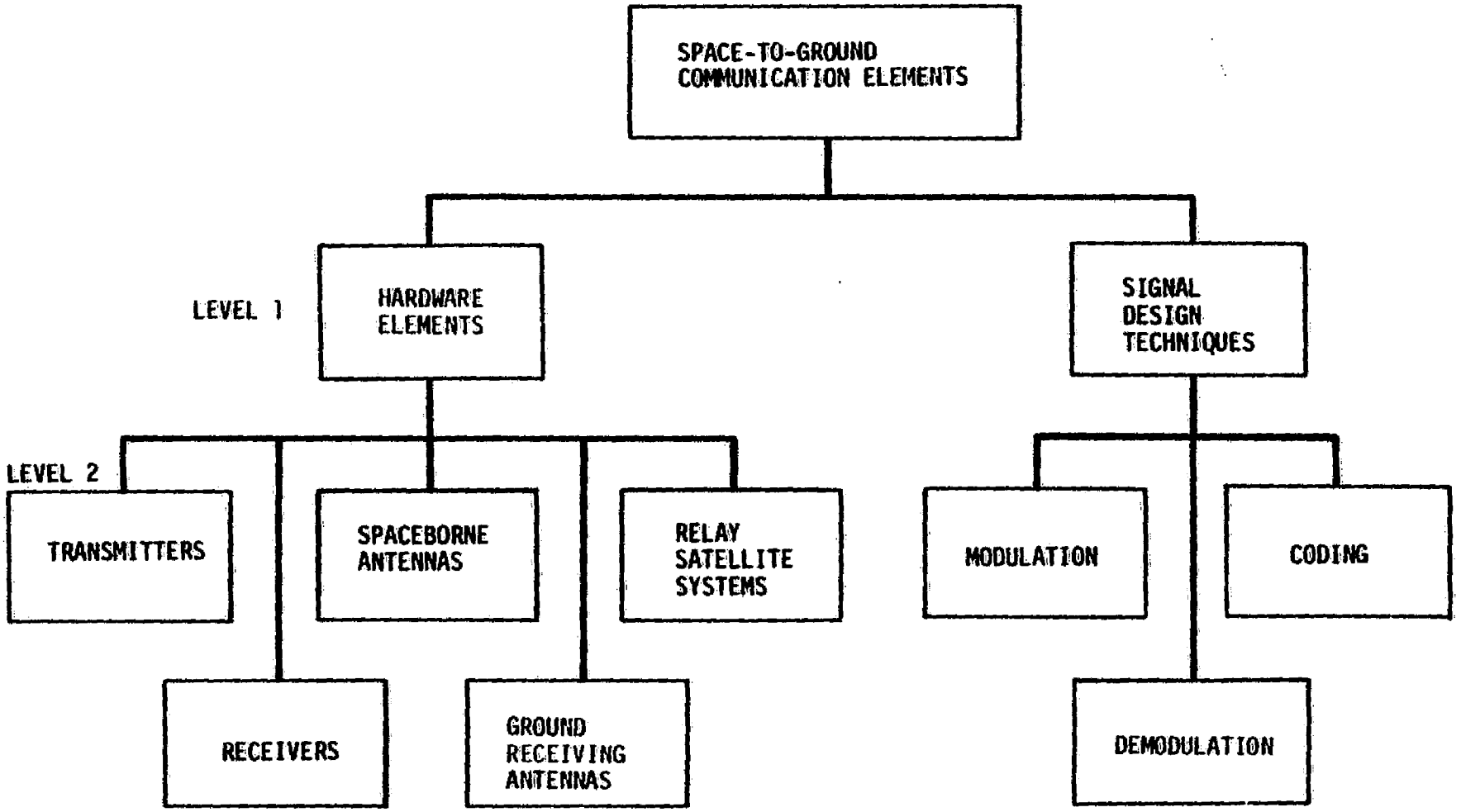


FIGURE 5-1. SPACE-TO-GROUND COMMUNICATION ELEMENTS

5.1 SPACEBORNE TRANSMITTERS

The trend in space communications is toward increased data rates, which dictates an increase in transmitter capability, including wider bandwidths and higher output power. In the event of extremely high data rates, transmitters must operate at higher frequency bands in order to accommodate the increased bandwidths. Future NASA programs, beginning with the space shuttle, will be making use of their higher operating frequencies.

The communications industry is involved in a number of major efforts to develop active signal-generating devices that are capable of improved performance at all assigned frequency bands. Although the major efforts appear to be going on at 35 GHz and below, developments are occurring at frequencies in excess of 200 GHz.

5.1.1 State of the Art in Spaceborne Transmitters

Frequency ranges and characteristics of interest for spaceborne transmitters are presented in Figure 5.1.1-1. This section presents state-of-the-art characteristics (power, efficiency, and gain) for the most important solid-state devices and Traveling Wave Tubes (TWTs) in the designated frequency bands, as well as lesser used devices such as klystrons, crossed field amplifiers (amplitrons and magnetrons), and backward wave oscillators. Pertinent data are presented in both tabular and graphic forms.

5.1.1.1 Solid-State Transmitter Devices - The most important solid-state transmitter devices in the frequency ranges of interest are bipolar transistors, gallium arsenide Field Effect Transistors (GaAs FETs), and IMPATT devices. Power, gain, and efficiency for state-of-the-art devices in each of these categories are presented in Tables 5.1.1.1-1 through 5.1.1.1-3. The data are presented as a function of frequency in the tables and in Figures 5.1.1.1-1 through 5.1.1.1-3.

Data in the tables and figures are for currently available production-type devices rather than devices that are available only in the laboratory. However, most of these solid-state devices are not space qualified since the advancement in technology has been so rapid.

Some of the major developments during the past year (1976-1977) have been in the area of GaAs FETs. Appendix A contains several tables listing specific state-of-the-art devices and their performance data. Efficiencies of the lower frequency GaAs FET devices are noteworthy, particularly one laboratory-type device that is 68% efficient at 4 GHz. Typically, GaAs FETs offer high efficiency, high gain, and medium power with respect to other solid-state devices. These devices are currently approaching the effectiveness of bipolar transistor devices in the 4- to 6-GHz frequency range. GaAs FETs also offer low noise and good efficiency in the 6- to 13-GHz range, with power levels ranging from 4 W at lower X-band to 1 W at upper X-band. Above this point, IMPATT diodes provide high power and moderate gain and efficiency. GaAs FETs are replacing

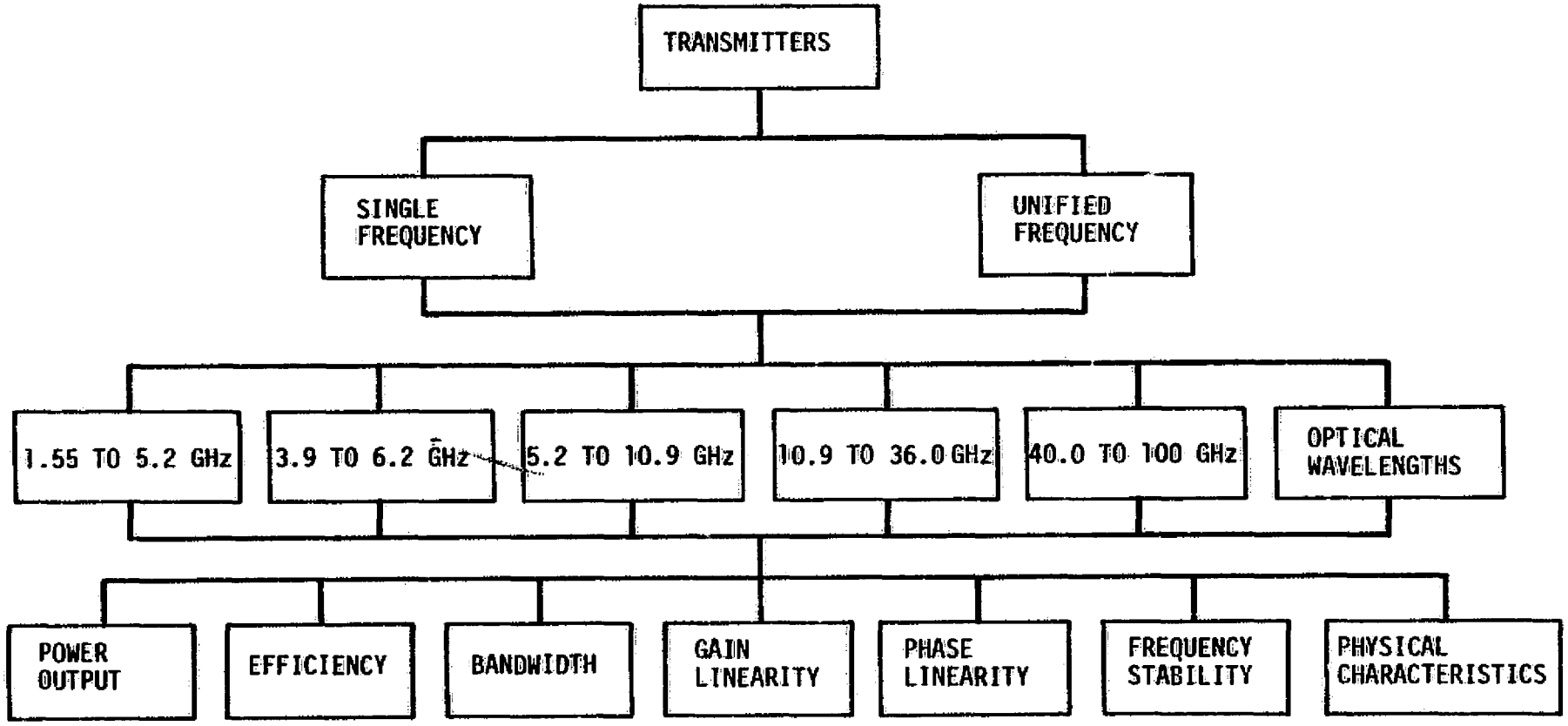


FIGURE 5.1.1-1. SPACEBORNE TRANSMITTER PARAMETERS

TABLE 5.1.1.1-1. SUMMARY OF KEY CHARACTERISTICS FOR STATE-OF-THE-ART BIPOLAR TRANSISTORS

CHARACTERISTIC	FREQUENCY (GHz)			
	1	4	6	10
Power Output (W)	40	6	2.5	0.7
Gain (dB)	9	6.5	5.5	3.7
Efficiency (%)	65	37	31	25

TABLE 5.1.1.1-2. SUMMARY OF KEY CHARACTERISTICS FOR STATE-OF-THE-ART GaAs FET DEVICES

CHARACTERISTIC	FREQUENCY (GHz)			
	4	6	10	40
Power Output (W)	3	1.5	1	0.1
Gain (dB)	11	10	8.5	2.5
Efficiency (%)	45	30	18	1

TABLE 5.1.1.1-3. SUMMARY OF KEY CHARACTERISTICS FOR STATE-OF-THE-ART IMPATT DEVICES

CHARACTERISTIC	FREQUENCY (GHz)			
	6	10	40	100
Power Output (W)	10	7	1	0.2
Gain (dB)	12	11	8	4.5
Efficiency (%)	40	30	10	4

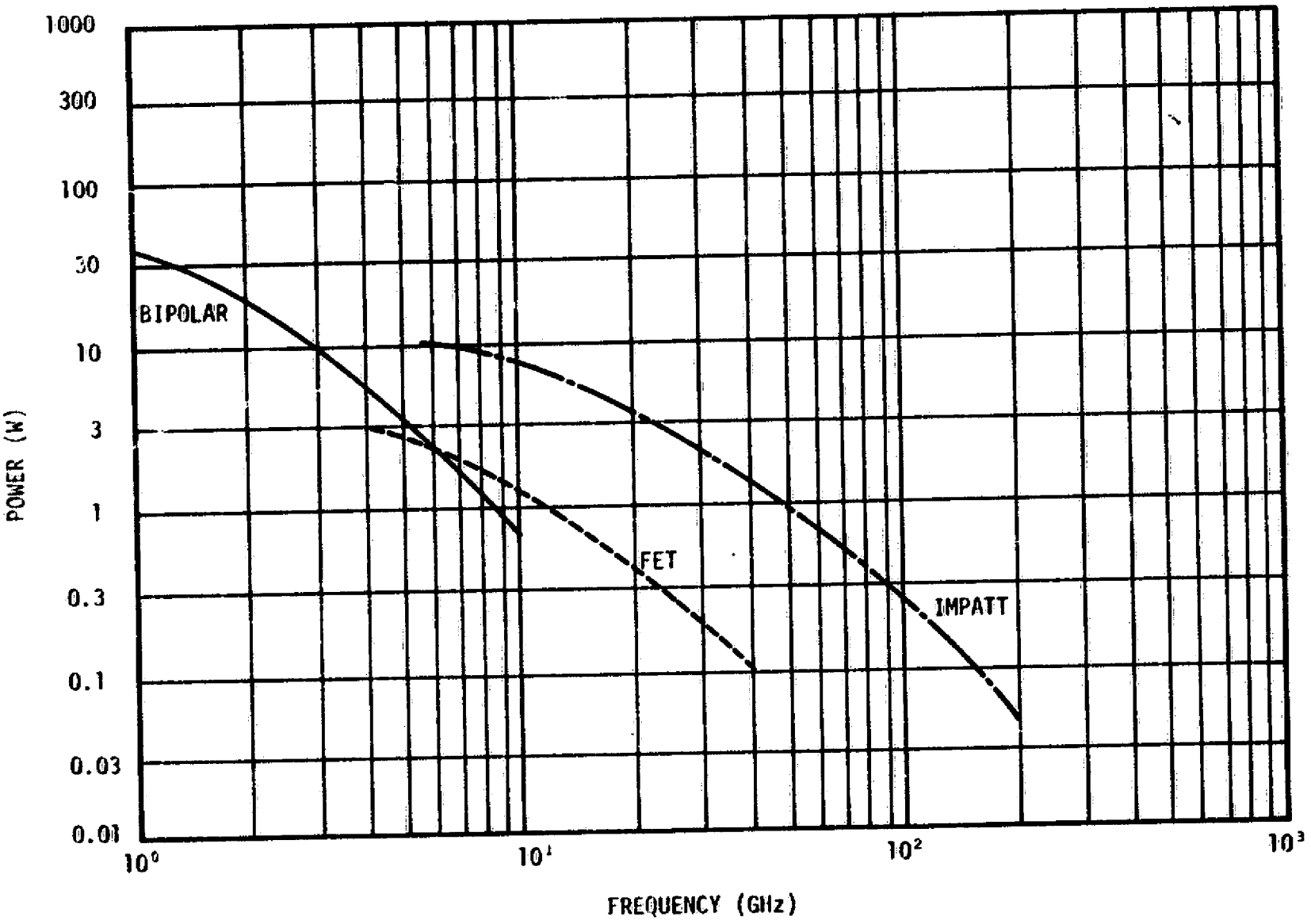


FIGURE 5.1.1.1-1. POWER AS A FUNCTION OF FREQUENCY FOR 1977 SOLID-STATE MICROWAVE DEVICES

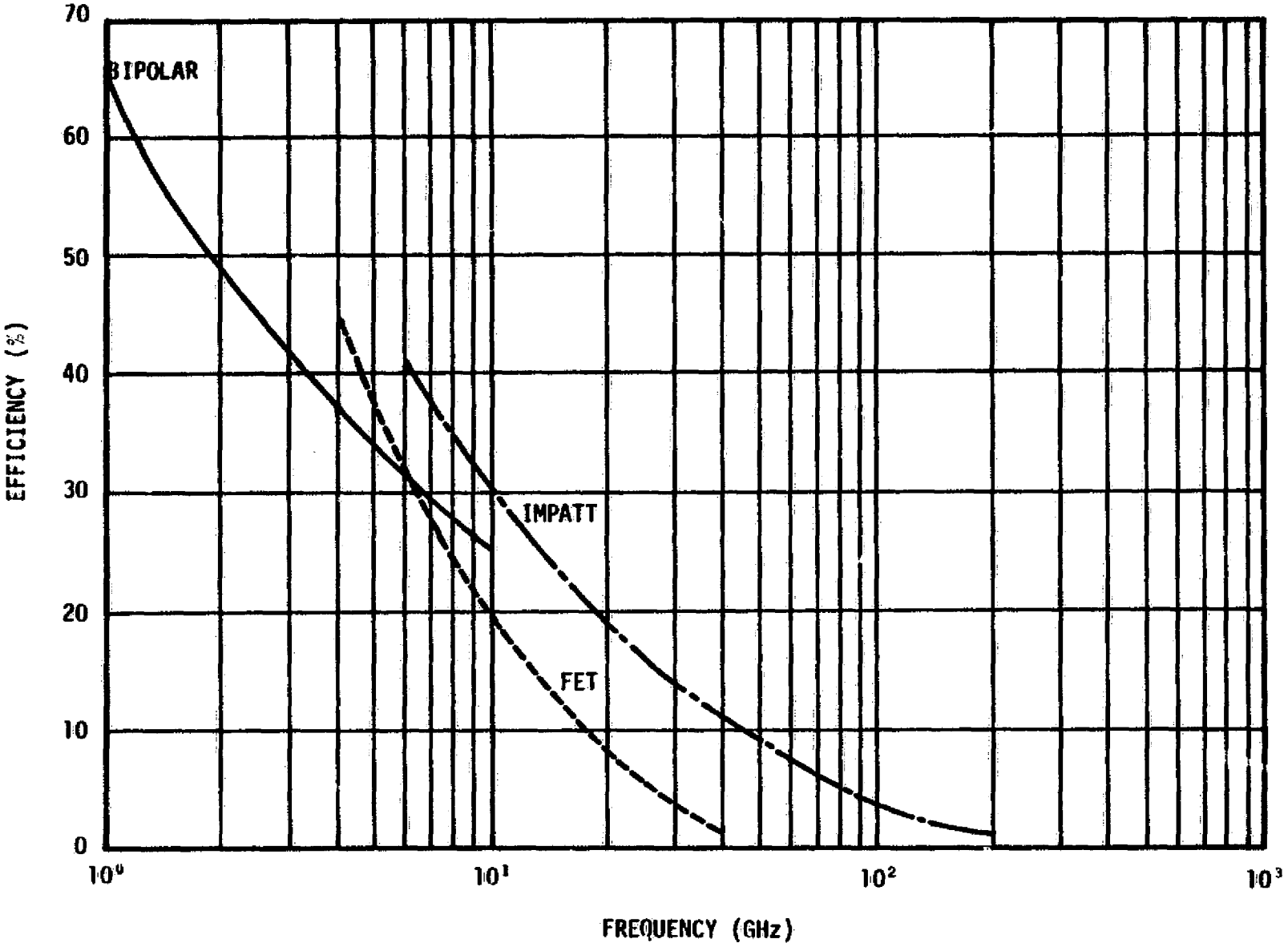


FIGURE 5.1.1.1-2. EFFICIENCY AS A FUNCTION OF FREQUENCY FOR 1977 SOLID-STATE MICROWAVE DEVICES

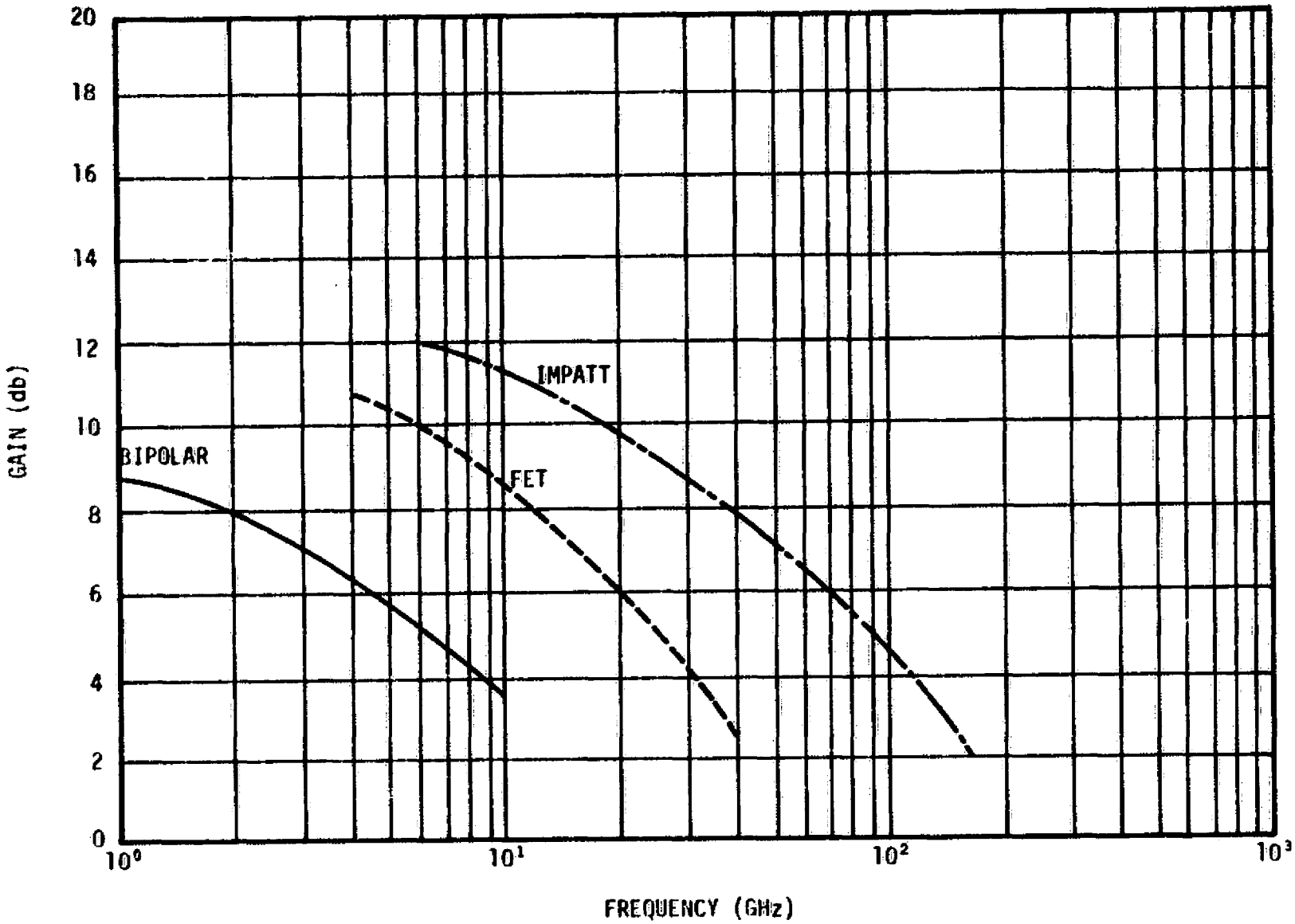


FIGURE 5.1.1.1-3. GAIN AS A FUNCTION OF FREQUENCY FOR 1977 SOLID-STATE MICROWAVE DEVICES

IMPATTs, low-noise TWTs, and Gunn diodes for power levels up to 1 W at frequencies below Ku-band. In addition to power amplifier applications, FETs offer a capability for microwave signal processing. Applications include the direct measurement of microwave signal frequencies and frequency synthesis using phase-locked loops and digital frequency dividers.

Other solid-state device technologies also are promising for microwave applications. One of the most promising is VMOS, a v-groove, n-channel, MOS field-effect device currently undergoing a major development effort. Three of the major semiconductor manufacturers either have, or will soon introduce, a line of RF power VMOS FETs. Devices are currently limited to 10 W at 400 MHz for production units, with laboratory models capable of operation at frequencies up to 700 MHz. One of the more favorable characteristics of VMOS is its very high switching speed (to 0.3 nsec). These speeds are useful for both microwave generation devices and direct microwave signal processing. Another technology that has potential use as a microwave device is double-diffusion MOS (DMOS) technology. Recent tests of the DMOS FET have demonstrated a 10-dB gain at 1 GHz, with greater than 40% efficiency. A 7-dB gain was obtained at 2 GHz, with usable gains extending to 12 GHz.

Table 5.1.1.1-4 presents the characteristics for several state-of-the-art solid-state amplifiers operating in the 9- to 11-GHz frequency range. These devices use IMPATTs, FETs, and a hybrid combination of the two. The FET amplifier has better efficiency, although the power output is the lowest. The noise figure for one of the FET stages is about 20 dB lower than for an IMPATT stage. The overall amplifier noise figure is 3 to 4 dB lower for the all-FET amplifier than for the hybrid amplifier. The highest gain is obtained with the hydrid amplifier, being some 3 to 4 dB higher than the other units.

The microwave power combiner provides a method of obtaining higher output power from a solid-state amplifier. Typical broadband combiners employ a central cavity with radial transmission lines leading out to the active devices. Each device is driven in phase, and the outputs are added to achieve power levels above that of an individual device.

TABLE 5.1.1.1-4. STATE OF THE ART FOR 9- TO 11-GHz POWER AMPLIFIERS

CHARACTERISTIC* \ ACTIVE DEVICE	IMPATT	IMPATT	FET-IMPATT HYBRID	FET
FREQUENCY (GHz)	9.2 to 9.8	10.7 to 11.7	9.1 to 9.8	9.2 to 9.8
POWER (W)	3	2	1.2	1
EFFICIENCY (%)	8	2	8	13
BANDWIDTH (MHz)	600 (1 dB)	300 (1 dB)	700 (3 dB)	600 (1 dB)
GAIN (dB)	22	23	26	22

*Based on actual state-of-the-art devices

Since the output is the sum of several devices, a failure of any device results in a graceful degradation of the power rather than a catastrophic failure. Several of these devices exist for operation at X-band frequencies. A 12-way combiner using GaAs FETs achieves 4.4 W at 8.5 GHz with a combining efficiency of 87.4%. A similar technique uses a five-stage IMPATT amplifier having a peak pulse output power of 1,024 W at 9.9 GHz. The output consists of 64 diodes connected through a 32-way combiner. The amplifier achieves a 30-dB gain and has an overall efficiency of 20%.

5.1.1.2 Microwave Tubes - Microwave tubes continue to offer benefits in the space communications area despite the challenges from solid-state technology. The major types of microwave tubes include TWTs, klystrons, crossed-field tubes (amplitrons and magnetrons), and backward wave oscillators. The most important microwave tube by far is the TWT, since it can be used as a high-power output stage. Second to TWTs is the klystron, a tube useful as either a signal source (oscillator) or an amplifier.

5.1.1.2.1 Traveling Wave Tubes - TWTs have undergone significant improvements in power and efficiency during the past few years, especially in the 11- to 14-GHz satellite communications band. The SATCOM 12-GHz, 50% efficient, 200-W TWT developed by NASA is currently being used on

the Canadian Communications Technology Satellite (CTS). The technology used for this tube is currently being applied in the development of 100- and 200-W TWTs for operation in the 41- to 43-GHz and the 84- to 86-GHz satellite bands. The European Space Research Organization (ESRO) TWT (TH3525), a 12-GHz, 40% efficient, 20-W, 60-dB gain tube is also in operation on the CTS. Other space-qualified TWTs generally range from 25 to 30% efficiency for the 10- to 12-GHz band. Table 5.1.1.2.1-1 lists the major characteristics for several K-band tubes. Table 5.1.1.2.1-2

TABLE 5.1.1.2.1-1. SUMMARY OF KEY CHARACTERISTICS FOR STATE-OF-THE-ART K-BAND AIRBORNE TWTs

FREQUENCY (GHz)	POWER (W)	GAIN (dB)	EFFICIENCY (%)	APPLICATION
10 to 12	20	50	25 to 30	Domestic Satellites
12	200	TBD	50	CTS
12	20	60	40	CTS
35 to 40	12	28	20	TBD

TABLE 5.1.1.2.1-2. SUMMARY OF KEY CHARACTERISTICS FOR STATE-OF-THE-ART TWTs

CHARACTERISTIC	FREQUENCY (GHz)					
	1	4	6	10	40	100
Power Output (W)	50	100	100	200	12	100*
Gain (dB)	40	45	40	50	28	25*
Efficiency (%)	35	40	50	50	20	10*

*This tube is for ground use; all others are space-type tubes.

lists the characteristics for other state-of-the-art TWTs. All of the characteristics given in this table are for space-qualified tubes, except the unit operating at 100 GHz. In laboratory-type TWTs, the current major achievement is a coupled-cavity TWT with 900 W of continuous power output at a frequency of 94 GHz. The gain of this tube is 30 dB and the operating efficiency is 30%.

Helix-type TWTs are being used in the 35- to 40-GHz range. Watkins-Johnson has achieved 20% efficiency at 12 W, 35 to 40 GHz, with a gain of 28 ±1 dB in a high-CW-power-level TWT designed for satellite and airborne communications. A new variable-power TWT has been developed by Hughes Aircraft Company for use in Earth terminal satellite communications. The tube is capable of operating at power levels between 500 and 1,200 W, with an instantaneous bandwidth of 7.9 to 8.4 GHz (500 MHz).

TWTs are undergoing significant improvement in efficiency and size for the 5- to 200-W power range. The Air Force Avionics Laboratory at Wright-Patterson Air Force Base has developed a new cathode material that appears promising. The new cathode uses a pliable nickel-based material that is formed in the shape of a dish. Preliminary tests show the new material to have several advantages over conventional tungsten cathodes. The nickel alloy cathodes can be shaped to close tolerance easily, whereas the tungsten cathodes are extremely hard and must be machined into shape. The nickel also offers a lower operating temperature than is available with conventional cathodes. The new TWT, operating with a cathode current density of 10 A/cm^2, will reach a cathode temperature of 852°C, whereas the conventional cathode would reach a temperature of 1,000 to 1,100°C under the same operating conditions. One additional advantage of the new cathode is a shortened activation time. The new cathodes undergo an activation time of a few hours, whereas the old cathode material could require as long as several days for the process.

At frequency ranges below 12 GHz, the principal advantage of TWTs over solid-state devices is a higher efficiency (40 to 50%). Efficiency factors of 50% have been reached on several coupled-cavity tubes, and some research work indicates that 84% efficiency could be accomplished in a four-stage, coupled-cavity design.

5.1.1.2.2 Klystron Tubes - Klystrons have application in both transmitters and receivers. Klystrons can operate at frequencies as low as UHF and as high as a few hundred gigahertz. Table 5.1.1.2.2-1 presents frequency and power characteristics for reflex klystrons operating between 3 and 200 GHz. The individual efficiency of each tube is unavailable, but typical efficiencies range from 20 to 30%, degrading somewhat for the higher frequency tubes. Of special interest is the 10-mW tube operating in the 170- to 220-GHz range. This is a recently developed tube that represents the state of the art in millimeter tube technology, with 220 GHz appearing to be the frequency limit.

The two-cavity klystron oscillator is capable of a much greater power output than the reflex klystrons. Table 5.1.1.2.2-2 presents the characteristics for several typical two-cavity klystrons covering the 5-to 40-GHz range. Again, the individual efficiencies are unavailable, but the tubes typically have efficiencies from 20 to 30%. In the frequency range above 30 GHz, the extended-interaction klystron oscillator offers both good efficiency and relatively high power. Efficiencies for these devices run as high as 40%. A typical tube for the 30- to 50-GHz range has 100-W output, whereas the 50- to 80-GHz tubes are rated at 5 W. From 80 to 110 GHz, the typical power is 10 W, and 5 W can be obtained from the 110- to 140-GHz tube. These power ratings are for standard tubes. Similar devices with power output ratings of five to eight times higher can be obtained on special order.

5.1.1.2.3 Crossed-Field Tubes - Crossed-field tubes, such as magnetrons and amplitrons, are very efficient (over 80% in pulsed applications) at low frequencies (2 to 6 GHz) and high power (hundreds of watts at 6 GHz), but they are currently used for radar applications rather than communications. Work is in process at Raytheon to develop higher efficiency tubes and, at the same time, to lower the inherent noise. Testing on these new designs will include noise measurements, and it is thought that developments may bring about a design with applications for communications. Until this research is completed, the constraints of high noise and low gain (10 dB) limit these tubes in their usefulness for satellite communications.

TABLE 5.1.1.2.2-1. SUMMARY OF KEY CHARACTERISTICS FOR STATE-OF-THE-ART REFLEX KLYSTRONS

CHARACTERISTIC*	FREQUENCY (GHz)					
	3	5	10	50	100	200
Power Output (W)	1	2	1	0.35	0.3	0.01
Gain (dB)	20 to 30	20 to 30	20 to 30	20 to 30	TBD	TBD
Efficiency (%)	20 to 30	20 to 30	20 to 30	20 to 30	TBD	TBD

*For typical state-of-the-art production devices

TABLE 5.1.1.2.2-2. SUMMARY OF KEY CHARACTERISTICS FOR STATE-OF-THE-ART TWO-CAVITY KLYSTRONS

CHARACTERISTIC*	FREQUENCY (GHz)			
	5	10	20	40
Power Output (W)	5	10	4	1
Gain (dB)	30 to 40	30 to 40	30 to 40	30 to 40
Efficiency (%)	20 to 30	20 to 30	20 to 30	20 to 30

*For typical state-of-the-art production devices

5.1.2 Trends and Projected Developments in Spaceborne Transmitters

Transmitters will be using a variety of devices for different frequency bands in the mid 1980's. Table 5.1.2-1 presents a summary of the more common solid-state devices for use in the microwave region. This table lists the devices, their applications, and typical operating frequencies, in addition to future trends. The table indicates that many of the devices will be replaced by bipolar transistors, GaAs FETs, and, at the higher frequencies, by IMPATTs. Those special-purpose devices not being replaced will be capable of increased speed, power, and stability. Below 6 GHz, bipolar devices will be competing against TWTs for size, power, and efficiency. GaAs FETs will replace low-noise TWTs and bipolar devices above 5 GHz. The upper frequency limit of the GaAs FET will increase, and GaAs FETs will replace IMPATTs below 40 GHz. IMPATTs will continue to improve in both power and efficiency, reaching 2 W at 100 GHz by 1985. Tables 5.1.2-2 through 5.1.2-4 list the projected key characteristics for solid-state devices in 1980. Tables 5.1.2-5 through 5.1.2-7 present this information for 1985 devices. Power, efficiency, and gain projections based on the information in these tables are plotted as a function of frequency in Figures 5.1.2-1 through 5.1.2-3. Profiles for state-of-the-art devices are shown in the same figures for comparison.

Improvements in substrate technology are enabling greater power outputs and increased efficiency for FETs and IMPATTs. All major semiconductor manufacturers are striving to produce high-quality devices operating with state-of-the-art specifications. As in the past, the trend is for smaller package size, becoming 70 to 75% of the present size by 1980 and 45 to 55% by 1985. Costs of the devices will have a slight downward trend, with the 1985 costs becoming 90% of current cost for TWTs and 50 to 60% of the current cost for solid-state devices.

Other solid-state device technologies will continue to develop during the early 1980's. VMOS devices will develop into microwave devices as continued research allows increases in frequency. VMOS

TABLE 5.1.2-1. SUMMARY FOR SOLID-STATE MICROWAVE DEVICES

PRESENT DEVICES	APPLICATION	OPERATING FREQUENCIES	FUTURE TRENDS
Tunnel Diodes	Amplifiers, Oscillators, Detector/Mixers, Switching	0.1 to 40 GHz	Will be replaced by GaAs FETs
GaAs FET Transistors	Broadband Low-Noise Amplifiers, Tuned Oscillators, Receiver Front Ends	2 to 18 GHz	Devices will have higher power and will operate at higher frequencies; will replace low-noise TWTs and bipolar devices above 5 GHz; will compete with tunnel diodes for oscillators and mixers; will replace IMPATTs
Bipolar Transistors	Amplifiers, Oscillators, Low-Power Tube Replacement	VHF to 6 GHz	Lower cost in future; GaAs bipolar used for special applications and higher speed
Gunn Diodes	Oscillators, Low-Power Radar	2 to 60 GHz	Frequency will increase to 100 GHz; will replace multiplier step recovery diodes
TRAPATT Diodes	Microwave Oscillators, Low-Power Radar	1 to 10 GHz	Will be replaced by bipolar and GaAs FET devices

TABLE 5.1.2-1 - Concluded

PRESENT DEVICES	APPLICATION	OPERATING FREQUENCIES	FUTURE TRENDS
IMPATT Diodes	Microwave Oscillators, TWT Replacement	Above 5 GHz	Provide improved dc to RF efficiency for pulse and CW, increased use for millimeter wave oscillators; lower frequency devices being replaced by GaAs FETs
Schottky Barrier Diodes	Mixers, Detectors, Power Rectifiers	0.1 to 40 GHz	GaAs will replace silicon in these devices; frequency will be greater than 100 GHz; will replace point contact diodes
Parametric Diodes	Low-Noise, High-Frequency Amplifiers	2 to 40 GHz	Will be replaced by GaAs FETs below 10 GHz; increased use above 10 GHz
Pin Diodes	Phase Shifters for Phased Array Antennas, Switching, Attenuators, RF Limiters	Up to 20 GHz	Will provide higher power and faster response time
Varactor Diodes	Tunable Filters, Phase Shifters, Oscillators and Amplifiers	VHF to 35 GHz	Will achieve higher stability and higher switching speeds; GaAs will replace silicon

TABLE 5.1.2-2. PROJECTIONS OF KEY CHARACTERISTICS FOR BIPOLAR TRANSISTORS IN 1980

CHARACTERISTIC	FREQUENCY (GHz)			
	1	4	6	10
Power Output (W)	80	12	5.5	1.6
Gain (dB)	10	7.5	6.5	5
Efficiency (%)	67	40	34	29

TABLE 5.1.2-3. PROJECTIONS OF KEY CHARACTERISTICS FOR GaAs FET DEVICES IN 1980

CHARACTERISTIC	FREQUENCY (GHz)			
	4	6	10	40
Power Output (W)	7	4.5	2.5	0.3
Gain (dB)	12	11.5	10	4.5
Efficiency (%)	65	40	25	2.5

TABLE 5.1.2-4. PROJECTIONS OF KEY CHARACTERISTICS FOR IMPATT DEVICES IN 1980

CHARACTERISTIC	FREQUENCY (GHz)			
	6	10	40	100
Power Output (W)	20	15	3	0.7
Gain (dB)	13.5	12.5	9	6
Efficiency (%)	50	35	15	6

TABLE 5.1.2-5. PROJECTIONS OF KEY CHARACTERISTICS FOR BIPOLAR TRANSISTORS IN 1985

CHARACTERISTIC	FREQUENCY (GHz)			
	1	4	6	10
Power Output (W)	170	29	13	4
Gain (dB)	12	9.5	8.5	6.5
Efficiency (%)	70	44	39	34

TABLE 5.1.2-6. PROJECTIONS OF KEY CHARACTERISTICS FOR GaAs FET DEVICES IN 1985

CHARACTERISTIC	FREQUENCY (GHz)			
	4	6	10	40
Power Output (W)	10	8.5	5.5	1
Gain (dB)	13.5	13	12	7
Efficiency (%)	70	55	35	6

TABLE 5.1.2-7. PROJECTIONS OF KEY CHARACTERISTICS FOR IMPATT DEVICES IN 1985

CHARACTERISTIC	FREQUENCY (GHz)			
	6	10	40	100
Power Output (W)	50	40	8	1.5
Gain (dB)	16	15	11.5	8
Efficiency (%)	65	50	20	10

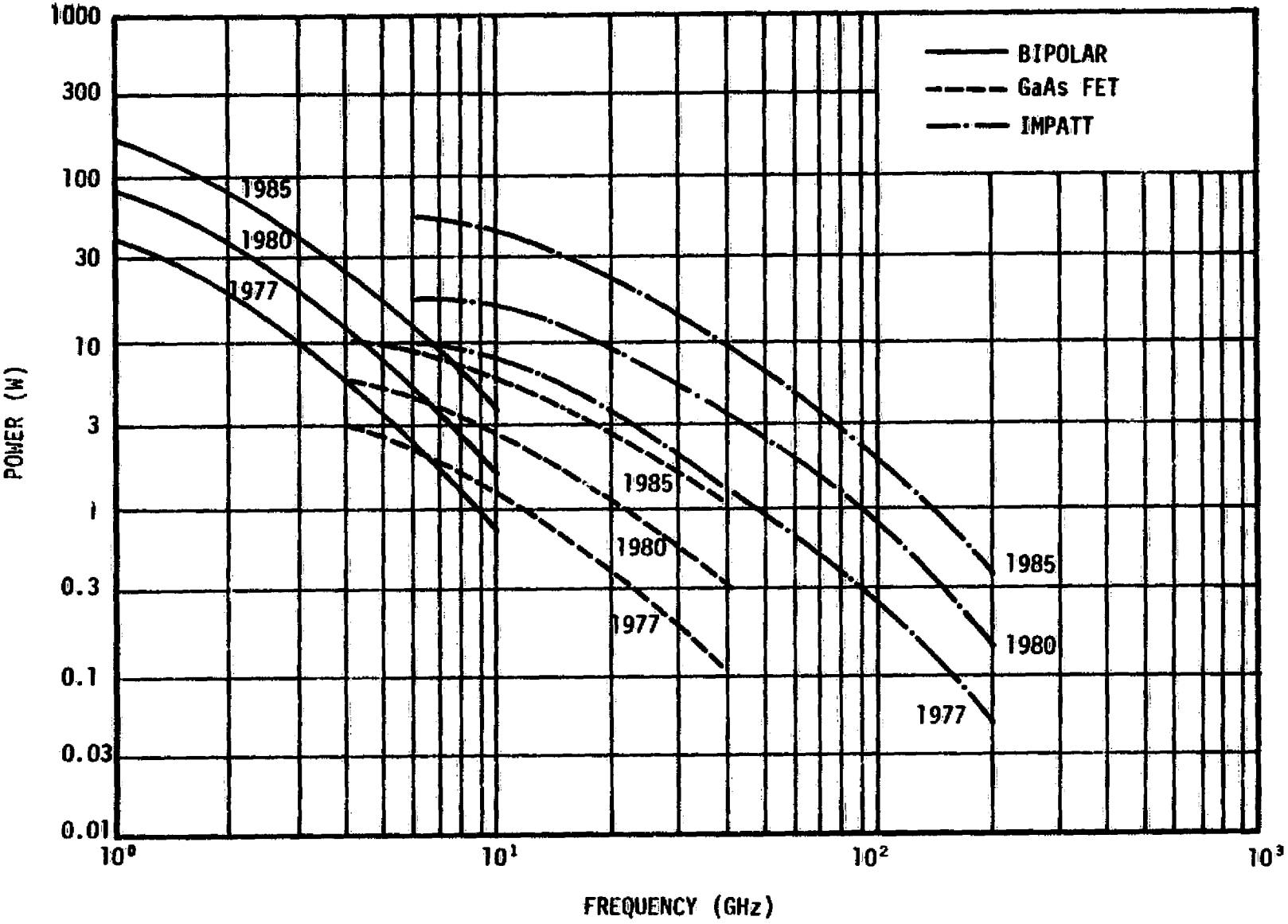


FIGURE 5.1.2-1. POWER AS A FUNCTION OF FREQUENCY FOR SOLID-STATE MICROWAVE DEVICES

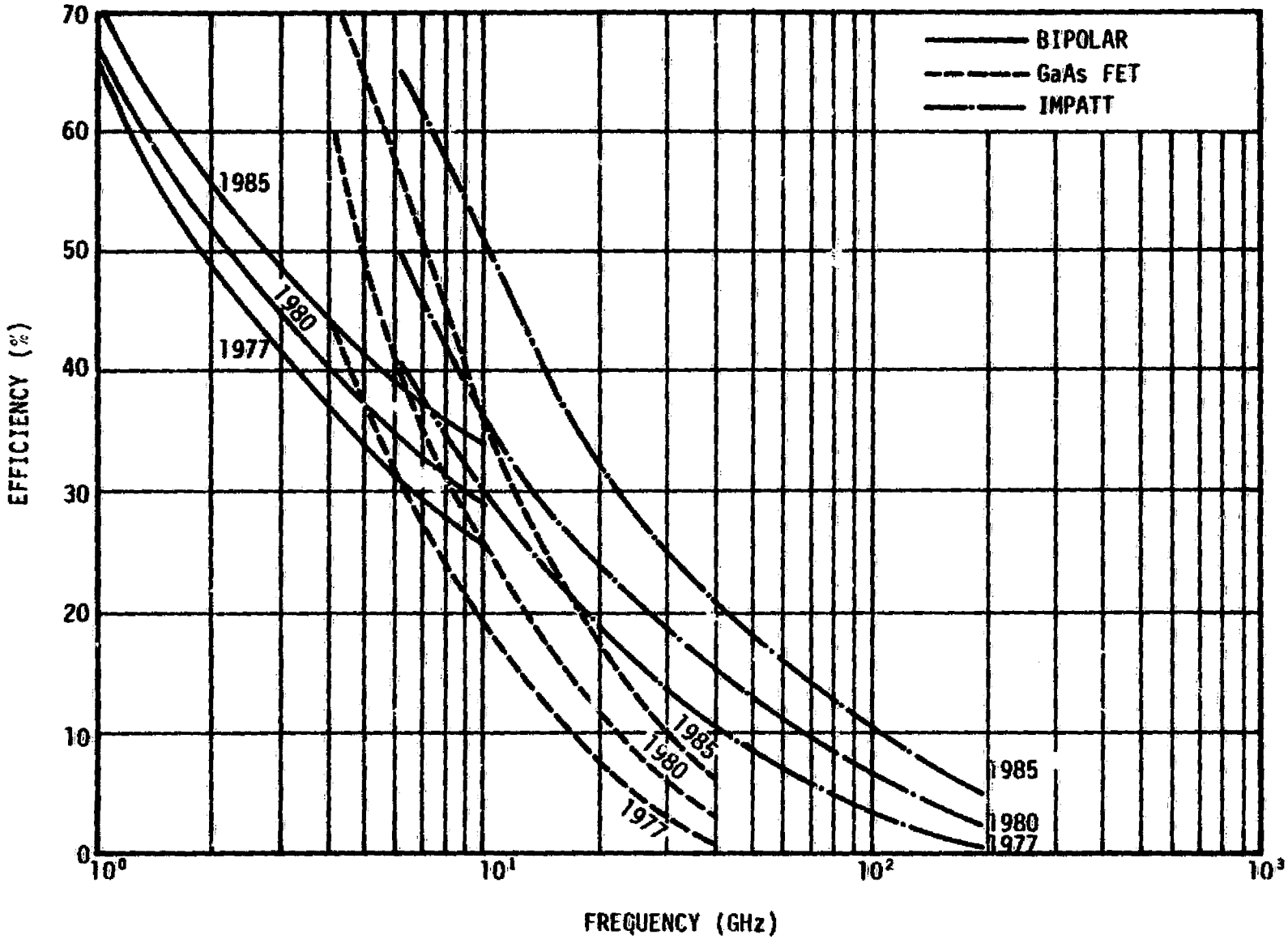


FIGURE 5.1.2-2. EFFICIENCY AS A FUNCTION OF FREQUENCY FOR SOLID-STATE MICROWAVE DEVICES

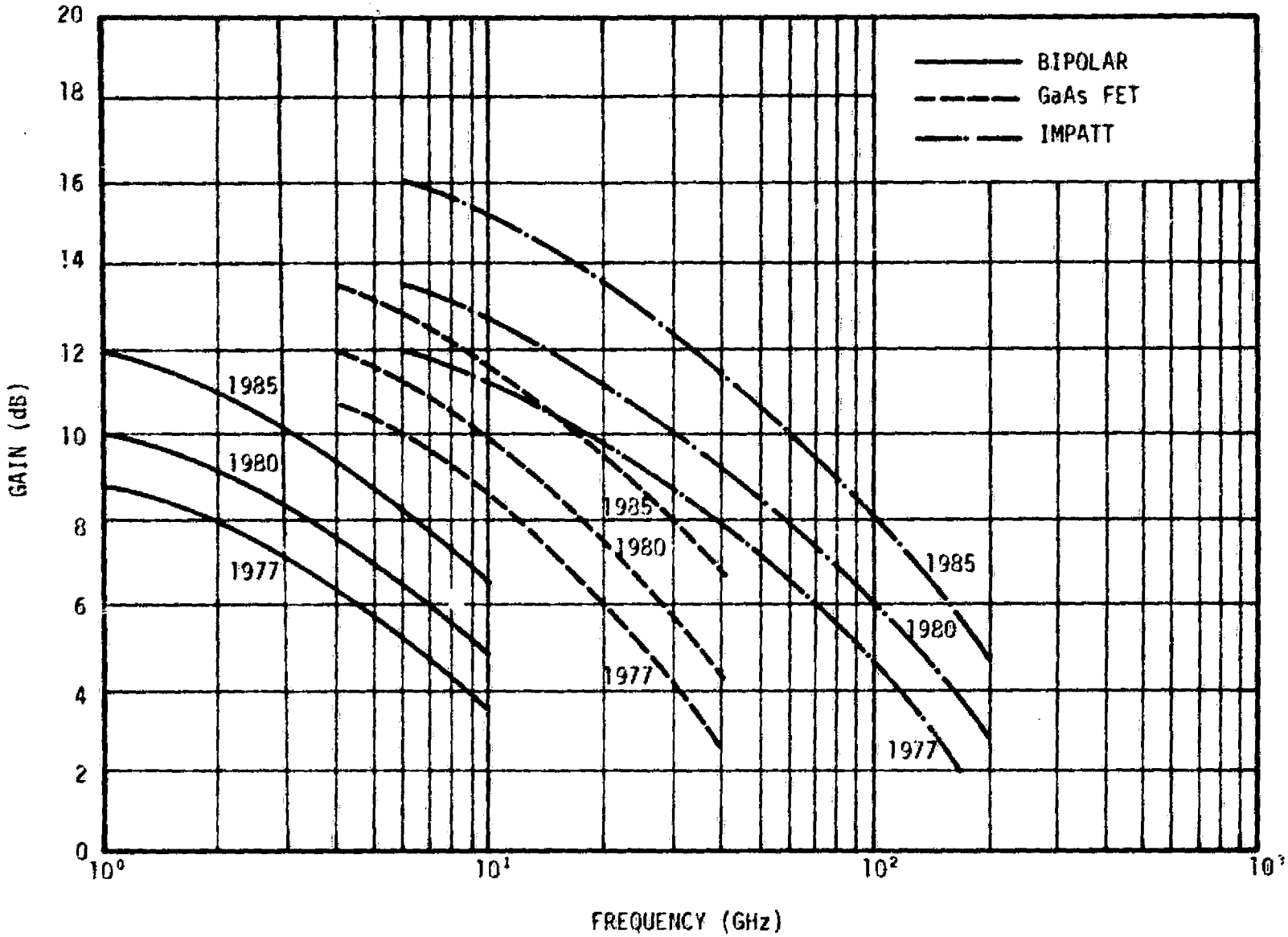


FIGURE 5.1.2-3. GAIN AS A FUNCTION OF FREQUENCY FOR SOLID-STATE MICROWAVE DEVICES

operation will be possible at S-band by the mid 1980's, giving a new source of low-noise, high-power amplifier devices. Additionally, DMOS operation at frequencies through Ku-band will be common. Other microwave device technology will develop as additional techniques of the high-speed computer logic industry are applied in power device areas.

A summary of the more common microwave tubes is given in Table 5.1.2-8. This table presents the tubes, their applications, typical operating frequencies, and future trends. The major microwave tube for the mid 1980's will continue to be the TWT. The trends for TWTs will include higher power and greater efficiency, especially at frequencies above 15 GHz. For low-power and low-noise applications, TWTs will be replaced by solid-state devices. For the higher power levels, TWTs will continue to be used. For wide-bandwidth applications (10% of operating frequency), helix-type TWTs will be used. For applications requiring only 1 to 2% bandwidth, the coupled-cavity tubes will offer increased power levels and greater efficiency. These higher-efficiency, narrow-bandwidth tubes will have their greatest impact at frequencies above 40 GHz. Tables 5.1.2-9 and 5.1.2-10 present the projected key characteristics for TWTs in 1980 and 1985. The data reflect greatest improvements in the 40- and 100-GHz ranges. This will occur as space-type tubes begin to approach the power and efficiency ratings of current terrestrial TWTs.

For other microwave tubes, the trend will be toward higher power and efficiencies also. High-power klystrons will operate at higher frequencies with increased efficiencies. The low-power klystrons will be replaced by solid-state devices. Crossed-field amplifiers will have increased power and efficiency and lower noise. Their communication-oriented applications will be limited to ground use, however.

Continued advancements in microwave devices depends heavily on demand as well as the research and development funds that are applied to these areas. Many of the devices covered in this section have a great demand in industrial uses as well as space applications. This is

TABLE 5.1.2-8. SUMMARY FOR MICROWAVE TUBES

PRESENT DEVICES	APPLICATION	OPERATING FREQUENCIES	FUTURE TRENDS
Traveling Wave Tubes	ECM, Radar, Avionics, Communications	1 to 100 GHz	For high-power TWTs: higher frequencies, greater power and efficiency For low-power TWTs: replacement by solid-state amplifiers below 10 GHz
Klystrons	Radar, ECM, Communications, Instrumentation	0.1 to 220 GHz	For higher-power tubes: gain and frequency will increase; solid-state replacement not likely. For low-power tubes: will be replaced by solid-state amplifiers
Crossed-Field Amplifiers	Radar Transmitters	1 to 17 GHz	Increased gain; peak power increased to 5 MW; solid-state replacement not likely
Backward Wave Oscillators	Type "O" Used for Microwave Test Equipment; Type "M" Used for ECM Equipment	0.5 to 350 GHz	Will be replaced by solid-state oscillators in the 1-to-12-GHz region
Magnetrons	Airborne Radars, ECM	0.5 to 40 GHz	For high-power tubes: increased frequency agility For low-power tubes: replacement by solid-state oscillators

TABLE 5.1.2-9. PROJECTIONS OF KEY CHARACTERISTICS FOR SPACEBORNE TWTs IN 1980

CHARACTERISTIC	FREQUENCY (GHz)					
	1	4	6	10	40	100
Power Output (W)	70	120	135	250	50	50
Gain (dB)	43	48	44	53	35	30
Efficiency (%)	40	45	55	55	25	18

TABLE 5.1.2-10. PROJECTIONS OF KEY CHARACTERISTICS FOR SPACEBORNE TWTs IN 1985

CHARACTERISTIC	FREQUENCY (GHz)					
	1	4	6	10	40	100
Power Output (W)	100	150	170	350	200	200
Gain (dB)	48	53	50	60	45	40
Efficiency (%)	45	52	60	60	35	30

especially true of the solid-state devices. For the higher-frequency TWTs, some industrial devices already exhibit very impressive characteristics. A gap exists here between the industry and space devices, and only some additional development efforts will close this gap. In general, the higher frequencies (above 30 GHz) and the optical wavelengths are receiving more development effort than the intermediate bands, the major cause for this being the allotment of higher frequencies to space-craft communication. Improvement of hardware at the 41- and 84-GHz bands should influence performance at the lower frequency bands also.

Projected trends in available transmitter power for different device types operating at different frequencies are given in Figures 5.1.2-4 through 5.1.2-9. These figures are based on the information contained in the tables giving state of the art and projections for the various microwave devices. These figures show trends of increasing power, with the largest gains coming for devices operating in the 40- to 100-GHz range. The values shown are for output stages using single devices. Higher power levels can be obtained with the solid-state devices by using a radial power combiner to efficiently add the output from a number of devices. This not only provides higher power levels but provides a graceful power degradation in the event of an output device failure. The general increase in device power is dependent on an increase in device efficiency. Trends in transmitter device efficiency are given in Figures 5.1.2-10 through 5.1.2-15. Efficiencies of the devices will increase in all cases, with the greatest increases below 40 GHz being for FETs and the greatest increases above 40 GHz being for TWTs. These figures reflect the device efficiencies. The efficiency of the last stage has an important role in determining overall transmitter efficiency. The system efficiency will generally run from one-quarter to one-half the efficiency of the final stage. The trends for transmitter device gain for several frequencies are shown in Figures 5.1.2-16 through 5.1.2-21. TWT gain is expected to increase at a greater rate than for solid-state devices. The TWT gain should reach 60 dB at 10 GHz. At frequencies above 40 GHz, the gain will increase to a level comparable with current TWTs operating at lower microwave frequencies.

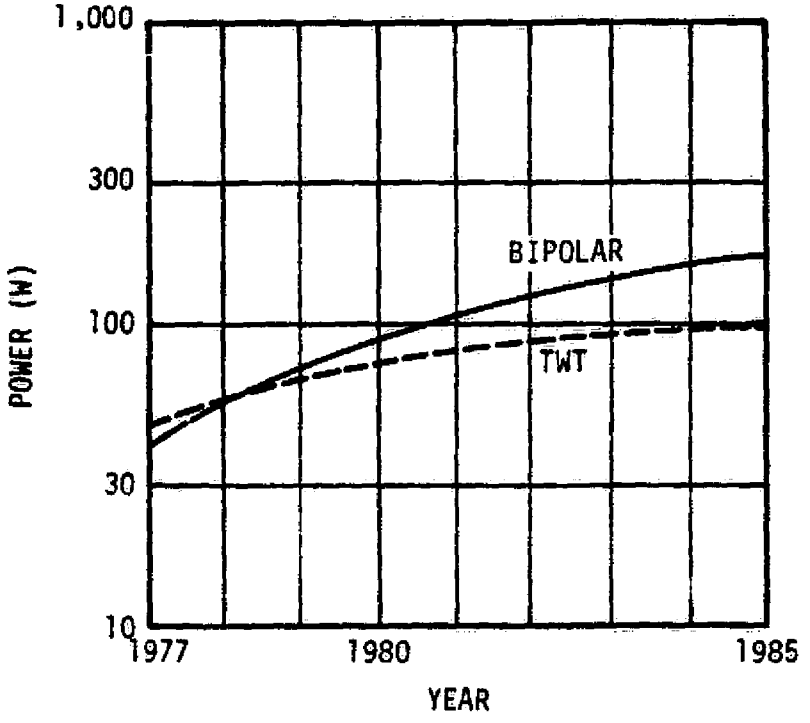


FIGURE 5.1.2-4. TRENDS IN TRANSMITTER POWER AT 1 GHz

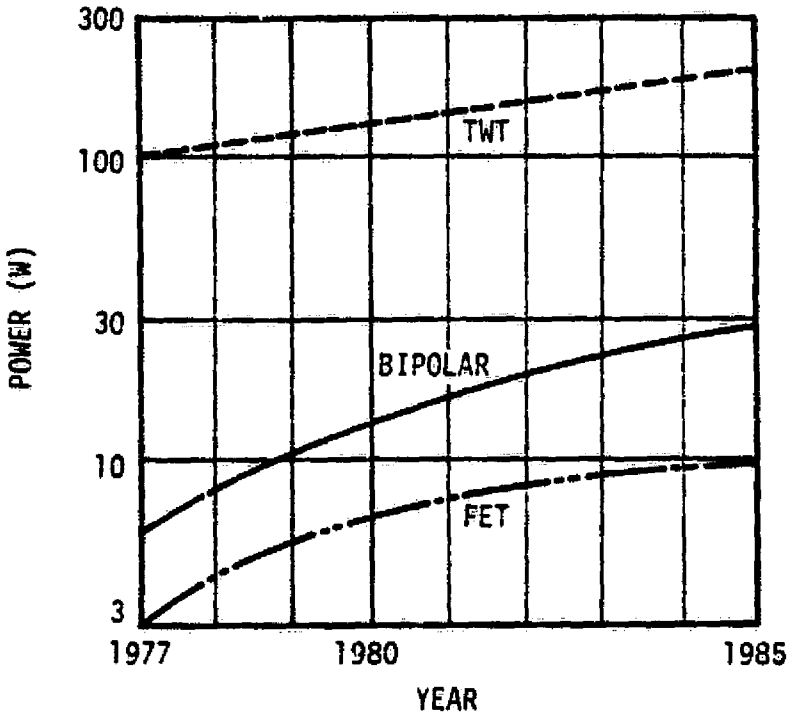


FIGURE 5.1.2-5. TRENDS IN TRANSMITTER POWER AT 4 GHz

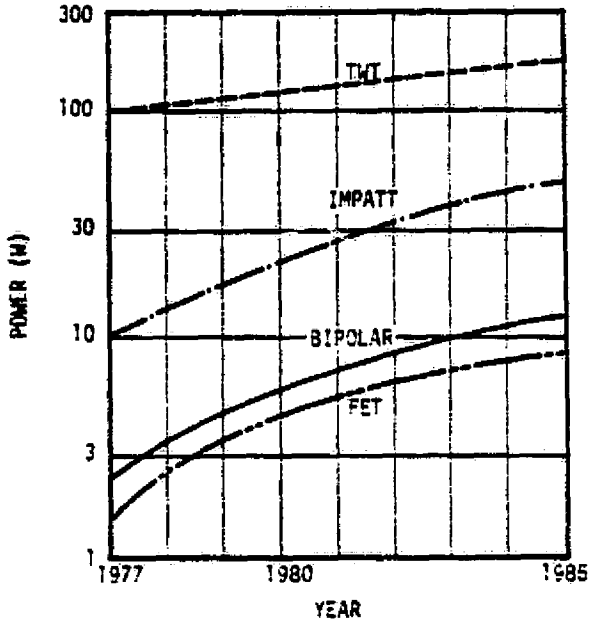


FIGURE 5.1.2-6. TRENDS IN TRANSMITTER POWER AT 6 GHz

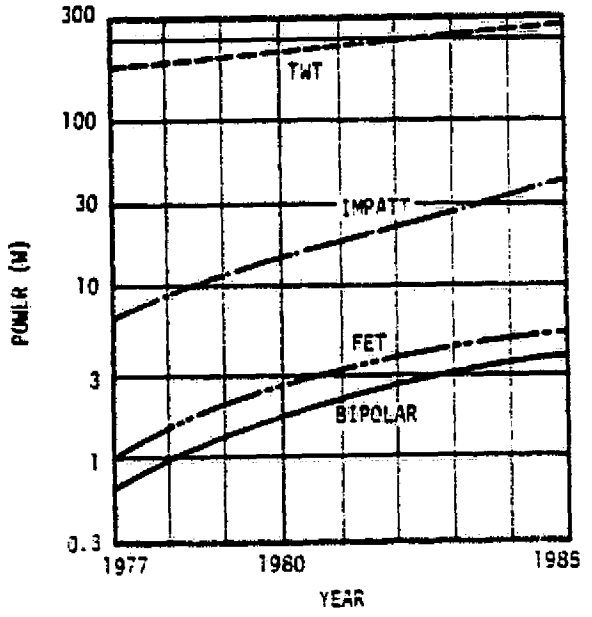


FIGURE 5.1.2-7. TRENDS IN TRANSMITTER POWER AT 10 GHz

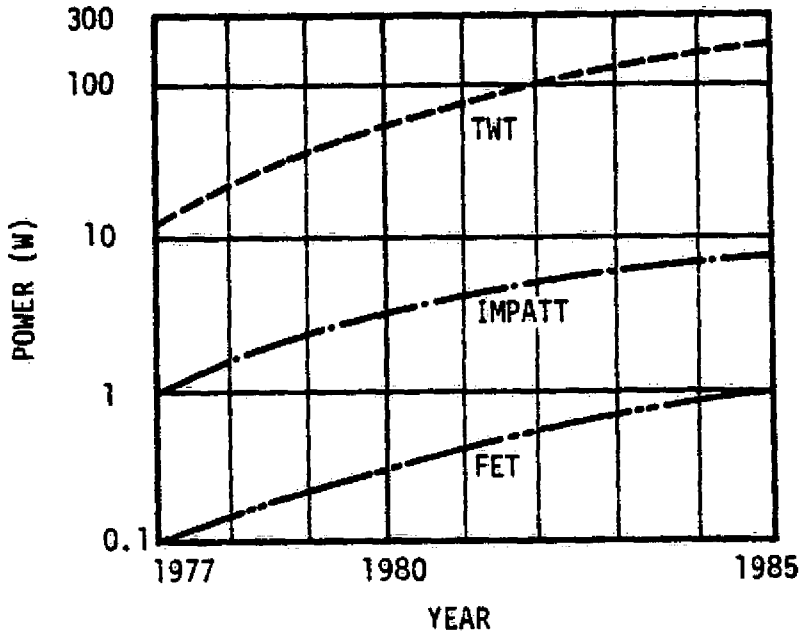


FIGURE 5.1.2-8. TRENDS IN TRANSMITTER POWER AT 40 GHz

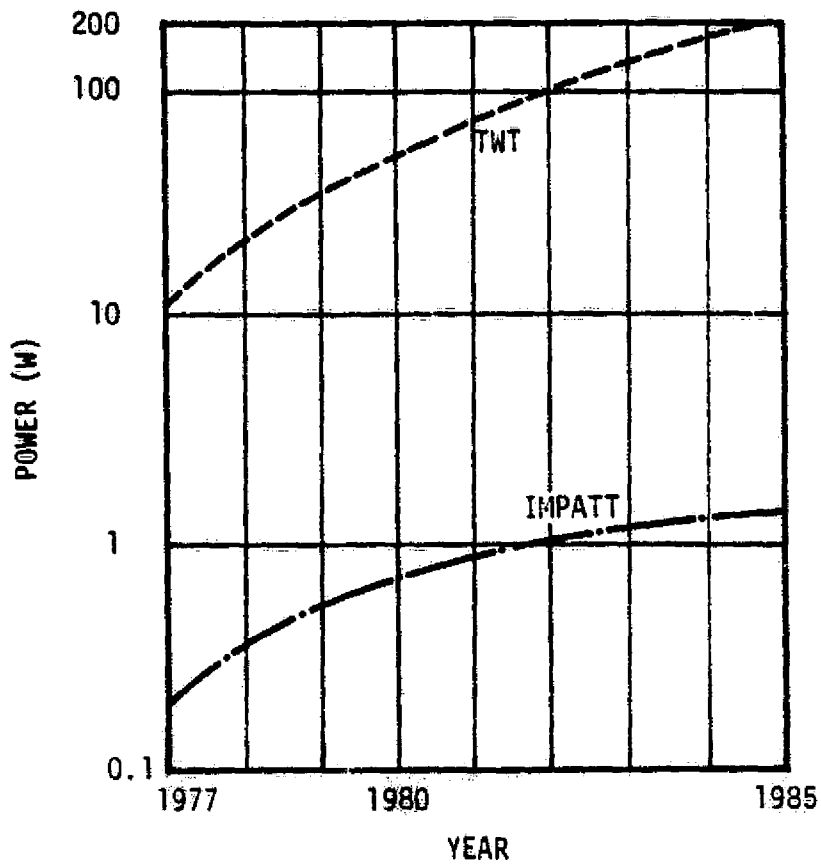


FIGURE 5.1.2-9. TRENDS IN TRANSMITTER POWER AT 100 GHz

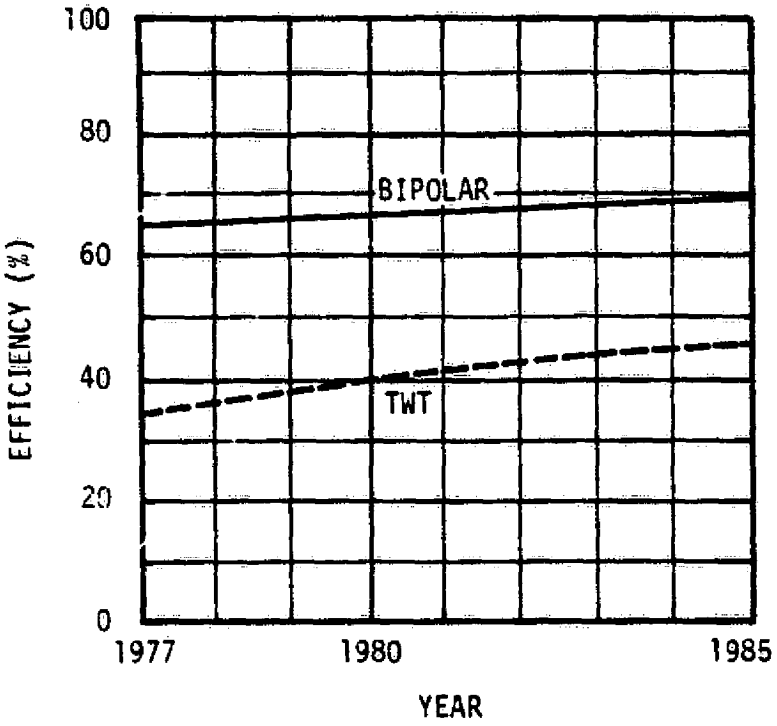


FIGURE 5.1.2-10. TRENDS IN TRANSMITTER DEVICE EFFICIENCY AT 1 GHz

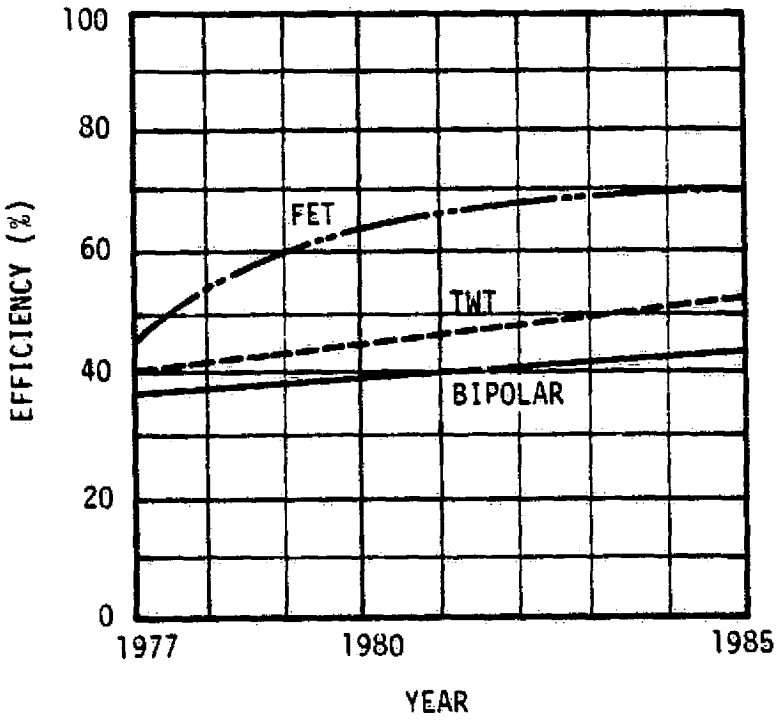


FIGURE 5.1.2-11. TRENDS IN TRANSMITTER DEVICE EFFICIENCY AT 4 GHz

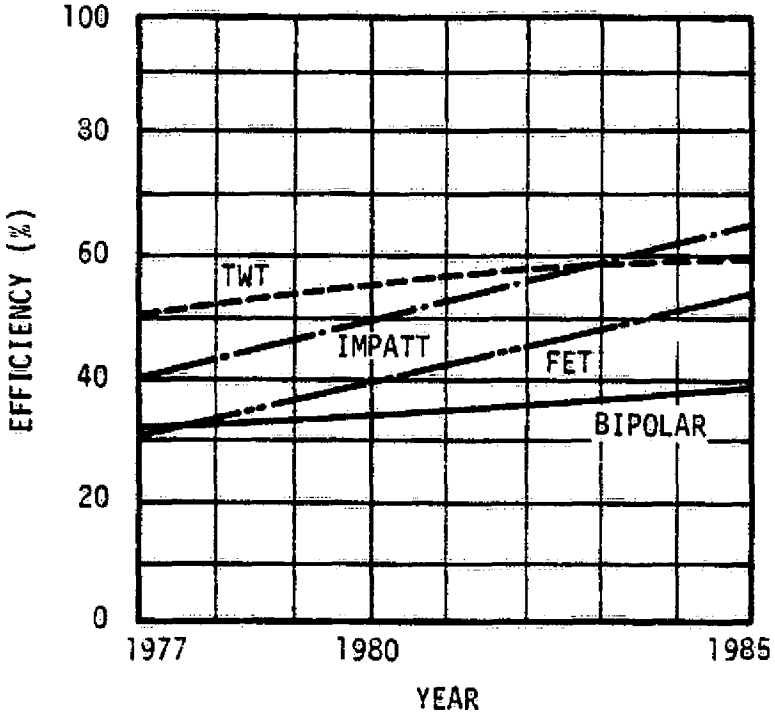


FIGURE 5.1.2-12. TRENDS IN TRANSMITTER DEVICE EFFICIENCY AT 6 GHz

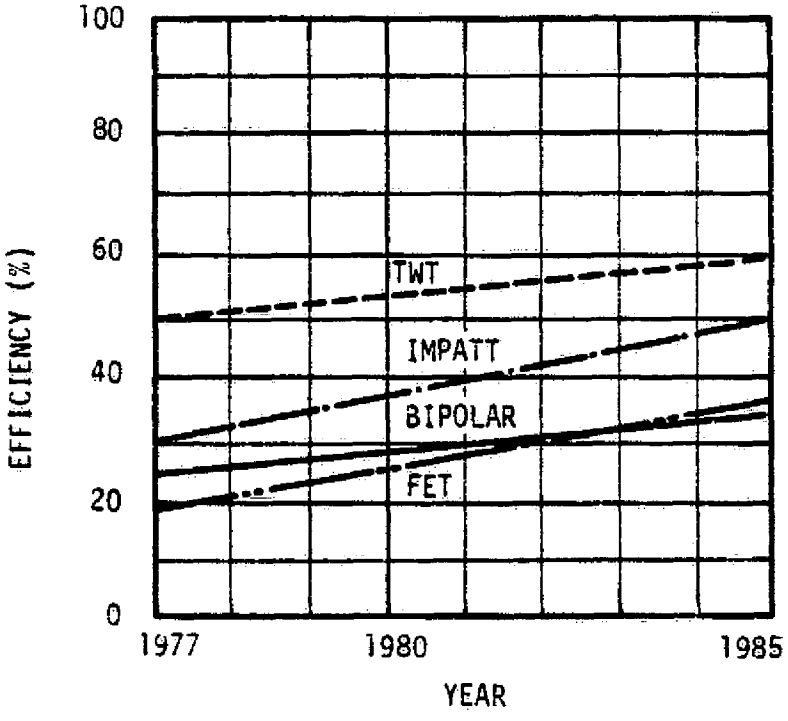


FIGURE 5.1.2-13. TRENDS IN TRANSMITTER DEVICE EFFICIENCY AT 10 GHz

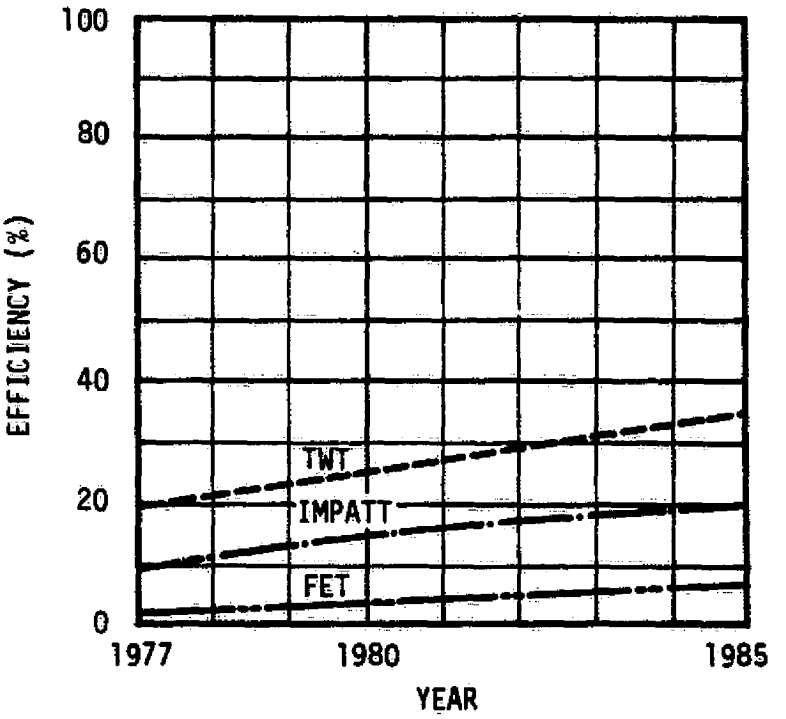


FIGURE 5.1.2-14. TRENDS IN TRANSMITTER DEVICE EFFICIENCY AT 40 GHz

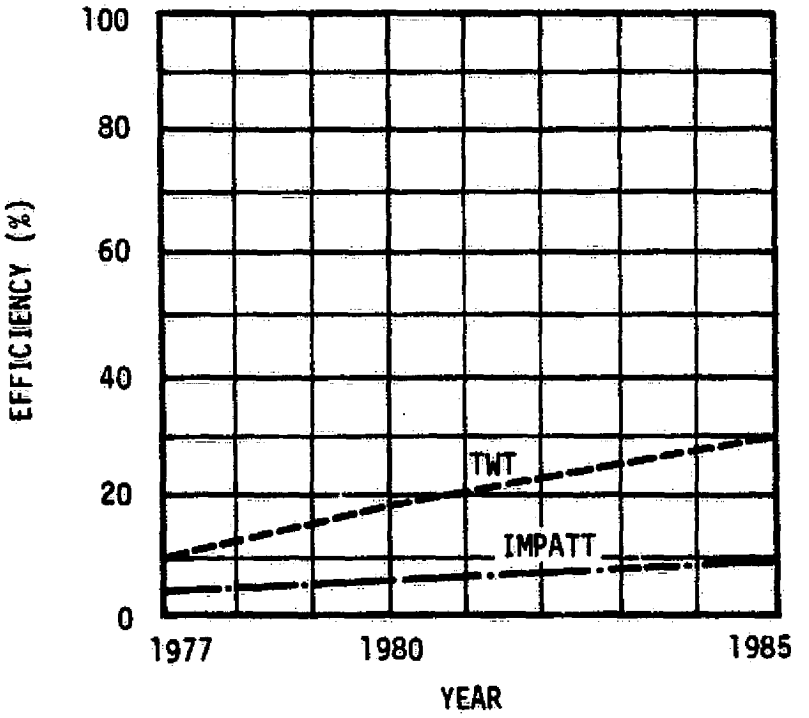


FIGURE 5.1.2-15. TRENDS IN TRANSMITTER DEVICE EFFICIENCY AT 100 GHz

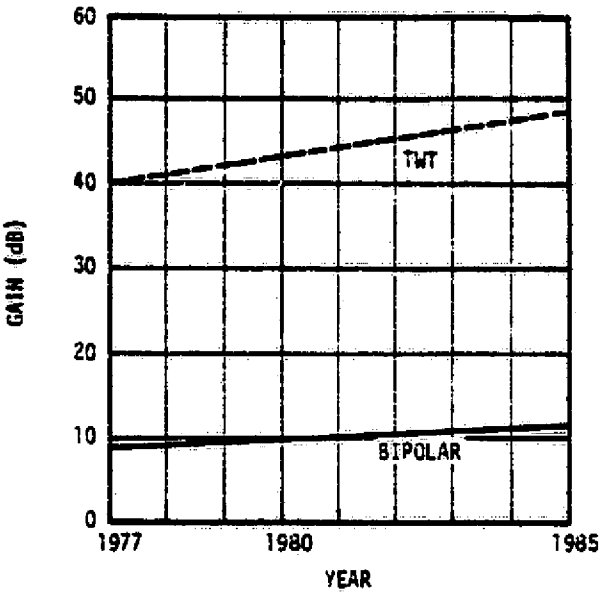


FIGURE 5.1.2-16. TRENDS IN TRANSMITTER DEVICE GAIN AT 1 GHz

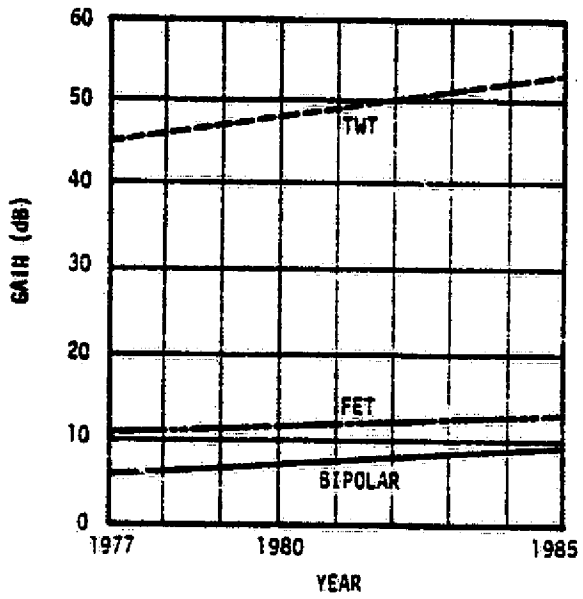


FIGURE 5.1.2-17. TRENDS IN TRANSMITTER DEVICE GAIN AT 4 GHz

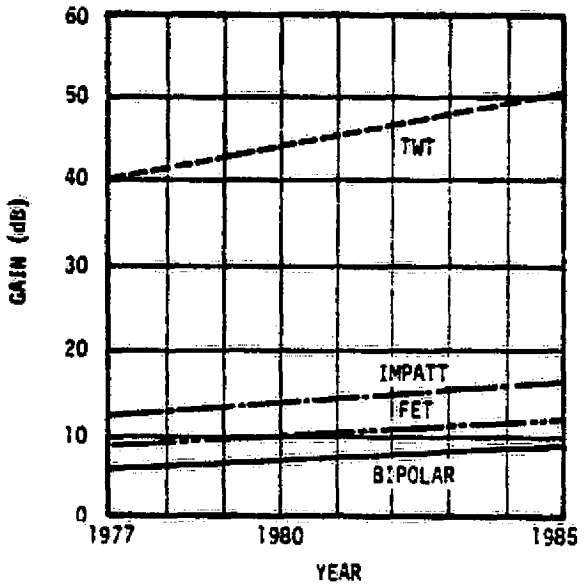


FIGURE 5.1.2-18. TRENDS IN TRANSMITTER DEVICE GAIN AT 6 GHz

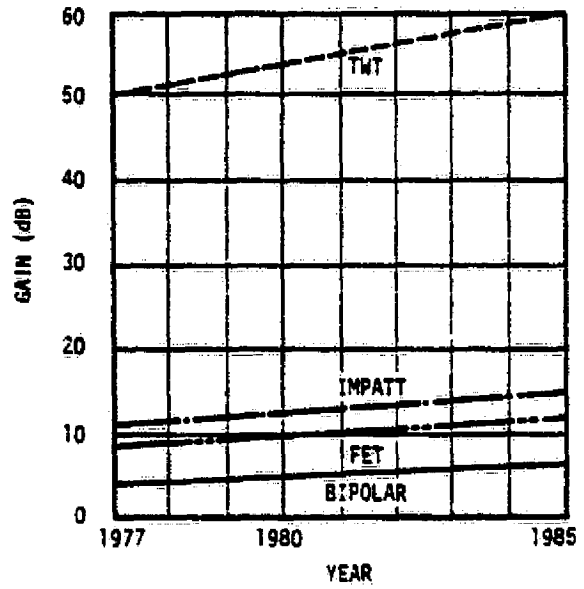


FIGURE 5.1.2-19. TRENDS IN TRANSMITTER DEVICE GAIN AT 10 GHz

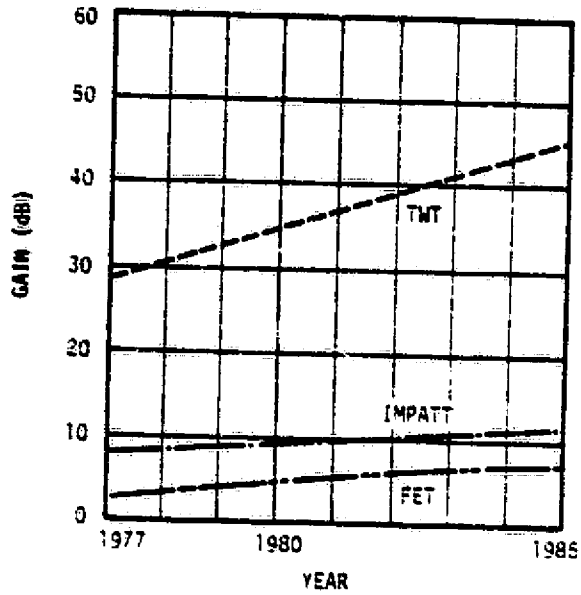


FIGURE 5.1.2-20. TRENDS IN TRANSMITTER DEVICE GAIN AT 40 GHz

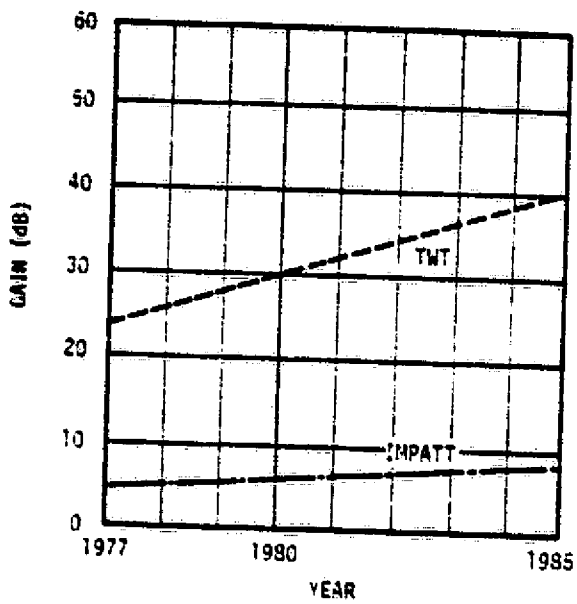


FIGURE 5.1.2-21. TRENDS IN TRANSMITTER DEVICE GAIN AT 100 GHz

5.2 SPACEBORNE ANTENNAS

Several categories of antennas are suitable for spaceborne communication applications. Figure 5.2-1 shows the types considered most important for future applications. The trend in the past has been for omnidirectional and low- and moderate-gain designs, but future uses will require higher gains and more involved designs. Space communication antennas currently operate on a number of frequency bands, but many of them are of minor importance. The two bands of major importance are S-band and Ku-band, since these will be the satellite communication bands in primary use in the 1980-1985 timeframe.

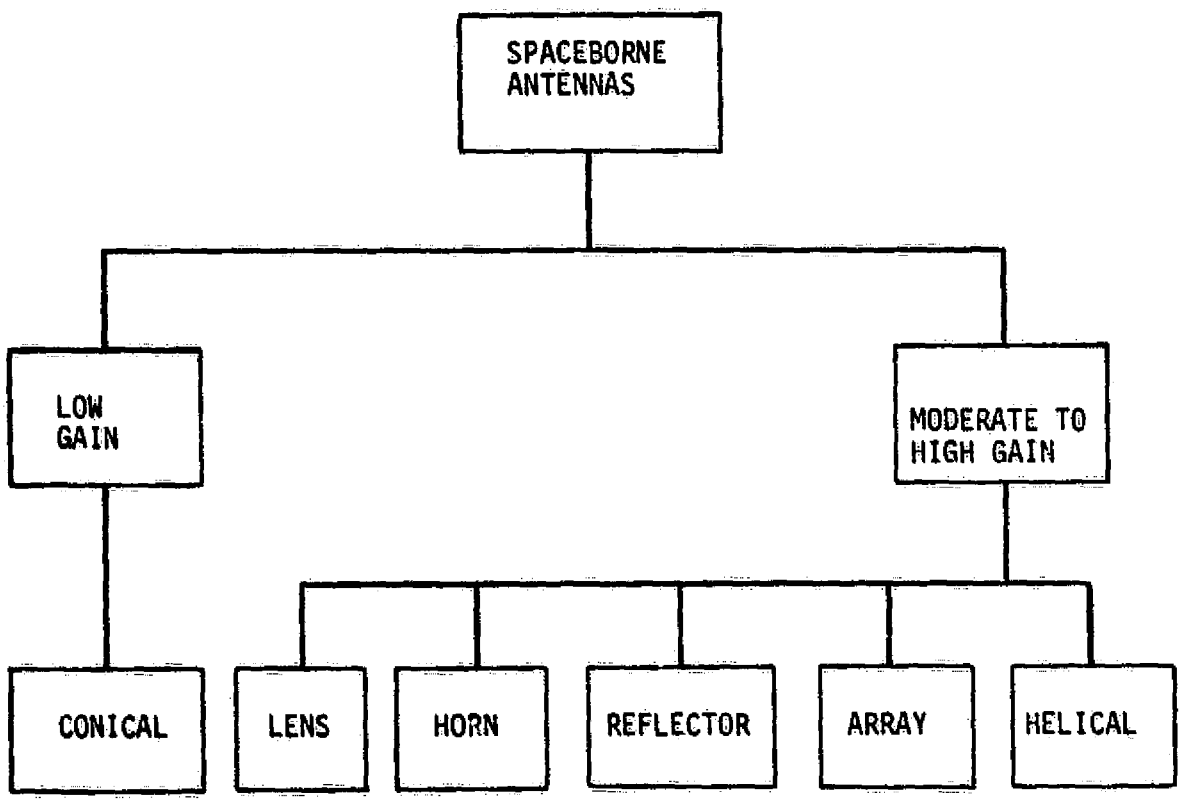


FIGURE 5.2-1. TYPES OF SPACEBORNE ANTENNAS

5.2.1 State of the Art in Spaceborne Communication Antennas

Most of the current spaceborne communication antennas operate at S-band. In the past, the downlink data rate has been low enough to allow the use of low-gain antennas for the low-Earth orbital operations. A common type of antenna for this application is the S-band conical antenna. This unit is physically small and can be deployed when necessary. It exhibits a 0-dB gain, a wide bandwidth, and a reasonably isotropic pattern. For applications with high data rates (above 50 Kbits/sec) or greater distances (synchronous and deep-space missions), it is necessary to use an antenna with gain. The types of antennas that have been most often used to achieve this gain are the reflector (dish) types and horn types. Some use has been made of helical antennas and fixed arrays employing simple, low-gain elements.

In view of increasing data rates and an increase in distance when using geosynchronous relay satellites, it has become necessary to develop high-gain spaceborne antennas capable of tracking. One such antenna is the Airborne Electronically Steerable Phased Array (AESPA) system being developed under the guidance of the Marshall Space Flight Center. The program has produced an S-band antenna system that has a diameter of 22 in. and a weight of 38 lb. The details of this 48-element antenna are contained in Table 5.2.1-1. In the transmit mode, it develops an Effective Isotropic Radiated Power (EIRP) of 35.8 dBW, and in the receive mode it has a boresight gain of 31.3 dB. The antenna array achieves a 5.6-dB noise figure in the receive mode. The phasing of the individual array elements is accomplished by a 3-bit, switched line arrangement. Work is currently underway on an AESPA system for Ku-band operation. The preliminary specifications for this system are given in Table 5.2.1-2. The proposed transmitting array uses 707 elements to achieve a 20-dB gain, while the receiving array uses 363 elements to obtain a 16-dB boresight gain. This system uses a ceramic microstrip 3-bit phase shifter to control each dipole element in the flat-plate array.

TABLE 5.2.1-1. S-BAND PHASED-ARRAY CHARACTERISTICS*

CHARACTERISTIC	SPECIFICATION
Coverage	Conical, ±60 deg from boresight
3-dB Beamwidth	16 deg
Overall size	22 in. diameter by 4 in.
Number of Elements	48
Polarization	RHCP
Weight	38.5 lb
dc Power	220 W
Transmitting Characteristics	
Frequency	2,252 to 2,312 MHz
Gain (boresight)	31.8 dB
Gain Loss at 60-deg Scan	4.4 dB
RF Drive Required	2.5 W
EIRP	35.8 dBW
Scan Angle (deg)	Sidelobe Level (dB)
0	-16
20	-11
40	-13
60	-8
Receiving Characteristics	
Frequency	2,071 to 2,131 MHz
Gain (boresight)	31.3 dB
Gain Loss at 60-deg Scan	4.0 dB
Noise Figure	5.6 dB
Scan Angle (deg)	Sidelobe Level (dB)
0	-15
20	-14
40	-12
60	-11

*This antenna system was developed under NASA's AESPA program.

TABLE 5.2.1-2. Ku-BAND PHASED ARRAY PRELIMINARY SPECIFICATIONS*

CHARACTERISTIC	SPECIFICATION
Coverage	Conical, ±60 deg from boresight
Size	Less than 1 m diameter by 10 cm thickness
Weight	Less than 25 lb
Radiating Elements	Active spirals or waveguide ports
Transmitting Characteristics	
Frequency	14.6 to 15.2 GHz
Gain (boresight)	Greater than 20 dB
Number of Elements	707
Receiving Characteristics	
Frequency	13.4 to 14.0 GHz
Gain (boresight)	Greater than 16 dB
Number of Elements	363

*This antenna system is being developed under NASA's AESPA Program.

Other work on airborne phased arrays includes a system to be operational in late 1980 on the Tracking and Data Relay Satellite System (TDRSS). The Adaptive Ground Implemented Phased Array (AGIPA) system will be used for the S-band multiple access service. The antenna is a 30-element multibeam steerable array capable of providing data relay and tracking services for up to 20 individual spacecraft simultaneously. The 20 beams (28 are theoretically possible) of this array will be controlled from the ground. The Air Force has several programs underway to develop an airborne X-band steerable phased array, but the details of these programs are currently unavailable. These units represent the

current state of the art in airborne phased-array antennas that have been built and tested. New-device technology in solid-state components and in microwave stripline waveguides will influence the development of other phased-array systems.

Another high-gain type antenna that is important in spacecraft applications is the reflector type antenna. Reflectors have been used in space for many years, but several new modifications have been used or proposed recently. One state-of-the-art reflector antenna is the deployable 30-ft parabolic dish used on the Applications Technology Satellite, ATS-6. This 132-lb antenna consists of three main sections: a hub, a set of 48 ribs, and a total of 48 reflective mesh panels. The unit is wrapped around the hub until ready for deployment, at which time it forms a 30-ft reflector. The antenna is capable of producing a 55-dB gain at a frequency of 8.25 GHz.

The majority of space reflector antennas used to date are of the "prime focus" type. Work is underway to develop dual-reflector (Cassegrainian or Gregorian) systems for space applications. This approach gives better on-axis gain than conventional single-reflector antennas. An example of one of the designs being studied is a 100-GHz Cassegrainian reflector. The system uses small corrugated feed horns offset so that no aperture blockage occurs. This allows a multiple-beam antenna with low sidelobe levels. This antenna has not been flown on any spacecraft, but it represents the type of work being done to develop the next generation of spaceborne antennas.

5.2.2 Trends in Spaceborne Antennas

The data from low-Earth-orbiting, unmanned spacecraft have in the past been transmitted directly to ground stations. This allowed the use of low-gain, omnidirectional antennas on the satellite. The satellite attitude was a major determinant as to the location and number of antennas used. The data rates were low (typically 50 Kbits/sec or less) by today's standards, relatively short distances were involved from low Earth orbit. The requirements for these spacecraft will change significantly by late 1980, however. With the operation of the TDRSS, all but three of the Space Tracking Data Network (STDN) ground stations will close. This forces spacecraft to relay data to Earth by way of geosynchronous satellites. As a result of the greater distances and higher data rates, the spacecraft will be required to use a high-gain tracking antenna system. For some spacecraft with special mission-dependent attitude requirements, the use of two or more of these antennas may be required. The data from future low-Earth-orbiting missions will fall into two categories: low data rates (50 Kbits/sec or less) and high data rates (multimegabit rates for imaging data). The low rates will be handled by the S-band Multiple Access (MA) service of TDRSS, and the high rate data will be handled by the S-band Single Access (SSA) service (up to 6 Mbits/sec) or the Ku-band Single Access (KSA) service (up to 300 Mbits/sec).

Requirements for these higher gains will result in the use of several types of high-gain antennas. Prime focus reflector type antennas will be deployable and in some cases collapsible. The use of dual-reflector antennas will become feasible as research work continues to produce lower sidelobe levels and as lighter materials are used in their design. The requirement will develop for steering systems as these high-gain, narrow-beamwidth antennas come into use. In addition to steerable reflectors, phased arrays will become more prevalent as advances are made in the state of the art of microminiaturization. New designs of phased arrays that use short-backfire antennas and other medium-gain elements to replace the current dipole designs will be developed. Increased use will be made of switchable-beam and multiple-beam antennas. New materials

will enable reductions in weight of many antenna designs, thus permitting the development of more complex antenna systems. The continuing development of feed systems will improve antenna gain performance and will reduce the sidelobe levels.

5.2.3 Projected Developments in Spaceborne Antennas

Reflector-type antennas will be the predominant design used through 1985. Reflectors will be used for all major communication bands, with prime focus reflectors being replaced somewhat by Cassegrainian reflectors. The use of multiple feeds for reflector antennas will become common. Tracking systems will be used with these reflectors to satisfy their pointing requirements (as in tracking the TDRSS). Folding-type reflectors will be common at frequencies below 10 GHz, and deployable design will be common for all communication bands. High-gain reflectors will be used through 1985 as the principal antenna type for frequency bands above 20 GHz.

During the 1980-1985 timeframe, phased-array antenna systems will become competitive with reflector types below 20 GHz. The phased-array systems will see increased performance and decreased weight and volume. Phased-array cost will decrease, reaching a point of being 1.5 to 2 times more expensive than steered reflector antennas. Phasing of groups of elements will become common as phase shifter designs allow the use of smaller phase increments. The dipole and slot antenna array elements will be replaced by more efficient, higher gain elements such as short-backfire elements. Arrays will handle much higher power levels without experiencing heat rejection problems or being subject to breakdown of the waveguides. Array feeds using low-loss air-stripline or microstrip networks with high permittivity substrates will allow designs with side-lobes that are down more than 30 dB. Advances in microprocessor controllers will allow adaptive designs capable of tracking and RF interference avoidance.

5.3 RELAY SATELLITE TECHNOLOGY FOR SPACE DATA TRANSFER

Relay satellites are addressed in this subsection from the standpoint of the technology and capability for relaying data from an orbiting spacecraft to the Earth. The capability is presented in terms of the orbiting spacecrafts' data throughput as a function of EIRP and other factors such as coding, single access by the spacecraft, and shared or multiple access with other users. The relay satellite technology for point-to-point and multipoint transfer is addressed in Section 9.1.2. Data presented herein are oriented more to planned capabilities than to technology, since all relay satellites that will be operational in the mid 1980's are already under development. Transmitter, receiver, and antenna technologies for low-Earth orbiting spacecraft and geosynchronous relay satellites are covered elsewhere in Section 5.

5.3.1 State of the Art in Space Data Transfer via Relay Satellite

Relay satellites capable of transferring wideband data from orbiting spacecraft on an operational basis are not currently available for use by civilian spacecraft. The TDRSS is currently under development for this purpose and is scheduled to be launched in mid 1980. The operational date for the satellite is November 15, 1980, with all but two or three STDN stations to be closed down beginning in early 1981.

The TDRSS network (to be provided by Western Union) will consist of two geostationary satellites located over the equator at approximately 41° and 171° W longitude, and a third in-orbit unit that serves as a spare. The Earth station for the system will be located in the continental United States in New Mexico. The satellites are positioned so as to permit direct contact with low-Earth-orbiting satellites for approximately 85% of the time. The coverage exclusion zone, centered over the Indian Ocean, is a function of the user spacecraft orbit inclination and altitude. For example, an orbital altitude of 200 km permits coverage for at least 85% of the time. An altitude between 1,200 and 2,000 km will receive 100% coverage, with the coverage decreasing toward zero as the satellite altitude approaches a synchronous orbit.

TDRSS capabilities are covered in a number of documents, including a document entitled "Tracking and Data Relay Satellite System (TDRSS) Users' Guide" (Ref. 5-1) by the Goddard Space Flight Center, from which the technical specifications provided herein were extracted. No other attempt will be made to repeat the information found elsewhere, other than to provide an overview of the primary technology parameters of bandwidth, sensitivity, power, etc., that affect the user interface with the spacecraft.

Each Tracking and Data Relay Satellite (TDRS) will provide multiple services to user spacecraft on S- and Ku-bands. Multiple Access (MA) service may be shared by up to 20 simultaneous users, each

with a data rate of 50 Kbits/sec or less. All MA users operate on the same frequency and are discriminated by unique Pseudo-random Noise (PN) codes. Two single-access services are available: Single Access at S-band (SSA) for data rates up to 6 Mbits/sec and Single-Access at Ku-band (KSA) for data rates of 300 Mbits/sec or less. The available RF bandwidths are 5 MHz on MA, 10 MHz on SSA, and 225 MHz on KSA, with coding requirements limiting the maximum return link data rates, especially on MA.

The link calculations for the MA return link (orbiting spacecraft to relay satellite to ground), the SSA return link, and the KSA return link have been extracted from the TDRSS Users' Guide and are included herein as Tables 5.3.1-1 through 5.3.1-3.

TABLE 5.3.1-1. SIGNAL CHARACTERISTICS FOR MA RETURN LINK, S-BAND

BER	10^{-5}
User EIRP	EIRP dBW
Space Loss	-192.2 dB
Polarization Loss	-1.0 dB
TDRS Antenna Gain at ±13 deg	28.0 dB
P_S at Output of Antenna	-165.2 dBW + EIRP
T_S (Antenna Output Terminals)	824 K
T_i (Due to Direct Other User Interference)	255 K
$K(T_S + T_i)$	-198.3 dBW/Hz
$P_S/K(T_S + T_i)$	+33.1 dB-Hz + EIRP
Transponder Loss	-2.0 dB
Demodulation Loss	-1.5 dB
PN Loss	-1.0 dB
Antenna Beam Forming Loss	-0.5 dB
System Margin	-3.0 dB
Required E_b/N_0 (ΔPSK)	-9.9 dB-Hz
Achievable Data Rate	+15.2 dB + EIRP
FEC Gain, R = 1/2, K = 7	5.2 dB
*Achievable Data Rate	+20.4 dB + EIRP

*This achievable data rate is the user's information rate. It should not be confused with the channel symbol rate, which is twice the information rate.

NOTE: Extracted from GSFC Document STDN No. 101.2 - TDRSS Users' Guide

TABLE 5.3.1-2. SIGNAL CHARACTERISTICS FOR SSA RETURN LINK

BER	10^{-5}
User EIRP	EIRP dBW
Space Loss	-192.2 dB
Pointing Loss	-0.5 dB
Polarization Loss	-0.5 dB
TDRS Antenna Gain	36.0 dB (50%)
P_S at Output of Antenna	-157.2 dBW + EIRP
T_S (Antenna Output Terminals)	586 K
KT_S at Output of Antenna	-200.9 dBW/Hz*
P_S/KT_S	43.7 dB-Hz + EIRP
Transponder Loss	-2.0 dB
Demodulation Loss	-1.5 dB
PN Loss	0.0 dB**
System Margin	-3.0 dB
Required E_b/N_O (ΔPSK)	-9.9 dB-Hz
Achievable Data Rate	27.3 dB + EIRP
FEC Gain, R = 1/2, K = 7	5.2 dB
***Achievable Data Rate	32.5 dB + EIRP

*Does not apply when operated simultaneously with DG2

**-1 dB for DG1

***This achievable data rate is the users' information rate. It should not be confused with the channel symbo rate, which is twice the information rate.

For more information, refer to TDRSS Users' Guide

NOTE: Extracted from GSFC Document STDN No. 101.2 - TDRSS Users' Guide

TABLE 5.3.1-3. SIGNAL CHARACTERISTICS FOR KSA RETURN LINK

BER	10^{-5}
User EIRP	EIRP dBW
Space Loss	-209.2 dB
Pointing Loss	-0.5 dB
Polarization Loss	-0.5 dB
TDRS Antenna Gain	52.6 dB (55%)
P_S at Output of Antenna	-157.6 dBW + EIRP
T_S (Antenna Output Terminals)	893 K
KT_S at Output of Antenna	-199.1 dBW/Hz*
P_S/KT_S	41.5 dB-Hz + EIRP
Transponder Loss	-2.0 dB
Demodulation Loss	-1.5 dB
PN Loss	0 dB**
System Margin	-3.0 dB
Required E_b/N_o (ΔPSK)	-9.9 dB-Hz
Achievable Data Rate	25.1 dB + EIRP
FEC Gain, R = 1/2, K = 7	5.2 dB
***Achievable Data Rate	30.3 dB + EIRP

*Does not apply when generated simultaneously with DG2

**-1 dB for DG1

***This achievable data rate is the users' information rate. It should not be confused with the channel symbol rate, which is twice the information rate.

For more information, refer to TDRSS Users' Guide

NOTE: Extracted from GSFC Document STDN No. 101.2 - TDRSS Users' Guide

A user currently receiving support by means of STDN is assumed to have the following equipment and capability:

- S-band receiver and transmitter
- S-band low-gain antenna
- Sufficient transmit EIRP and receiver sensitivity for operation with a 9-m ground antenna system.

MA users of the TDRSS must add equipment and capability as follows:

- A spread-spectrum receiver with PN correlator and a Phase-Locked Loop (PLL) capable of handling the frequency hop preamble used in acquisition
- A digital command decoder
- A spread spectrum transmitter
- Additional dc power for transmitter if continuous real-time operation is required
- A digital telemetry system capable of convolutional coding
- Increased EIRP, which may be accomplished by a higher gain antenna. Higher antenna gains and narrower beam-widths mean that the user will have to provide an antenna pointing (tracking and/or steering) capability.

SA users of the TDRSS must also have MA capability. SA requires higher receiver gains than MA, as well as higher EIRP.

The primary user interfaces on the ground will be the TDRSS ground terminal, the Goddard Space Flight Center (GSFC), and the Johnson Space Center (JSC). The communication link between these three ground centers will be implemented via the NASCOM network. The present configuration incorporates a link between the TDRSS ground station and GSFC, between the TDRSS ground station and JSC, and between JSC and GSFC. Each of these lines has a capacity of 1.544 Mbits/sec, but the users' data rate is limited to a nominal 1 Mbit/sec. The current NASCOM system configuration does not accommodate high multimegabit-rate data, but expansion of the network to accommodate up to 7 Mbits/sec is being considered. Other special cases may use leased services on a domestic communications satellite, with the receiver Earth terminal located at the user facility.

5.3.2 Trends in Space Data Transfer Via Relay Satellites

A number of trends in data relay satellites in general (migration to higher frequency bands, wider bandwidths, and higher data rates within these higher frequency bands) could affect space data transfer by means of relay satellites. The current TDRSS program (to be operational in late 1980) is defined for a 10-year lifetime, and there are no plans at this time to incorporate any of these changes. For a more detailed discussion of trends in data relay satellites in general, refer to Subsection 9.1.2.

The Air Force Avionics Laboratory, Wright-Patterson AFB, is working on a laser satellite relay experiment to be performed in 1979. This system will use an Nd:YAG laser to transmit data at 1,000 Mbits/sec from a geosynchronous satellite. NASA's Goddard Space Flight Center is developing an experimental laser data relay system called Advanced Tracking and Data Relay Satellite (ATDRS), which is similar in operation to the TDRSS. Their developed hardware has demonstrated a 300-Mbits/sec data rate and can track Doppler frequency variations over a ±700-MHz range. There are no current plans to implement this system.

For trends in relay satellite subsystems, refer to Subsections 5.1 and 5.2.

5.3.3 Projected Developments in Space Data Transfer Via Relay Satellites

Projections for the 1980-1985 timeframe indicate the use of the TDRSS for purposes of space data transfer by means of relay satellite. The current TDRSS program is defined to be operational until late 1990. Although an upgrading of the system would be technically feasible during the 1980-1985 timeframe, the current NASA planning does not incorporate any design changes.

The current TDRSS data rate is limited by allotted bandwidths rather than equipment limitations. However, data rates for the current NASA mission model are satisfied by TDRSS capabilities throughout the 1980-1985 timeframe. Since the TDRSS is a true repeater (bent-pipe) system, it is insensitive to the modulation type. In the event of a requirement for data rates greater than that provided by the system, a high-efficiency modulation technique (Subsection 5.6) could be implemented. The major restriction foreseen for the TDRSS system during its 10-year lifetime is a possible lack of bandwidth.

Projections for a second-generation space data relay satellite do not show implementation until 1990. Preliminary system studies are not expected to begin until 1982 or 1983, and serious design work on the system will not start until the 1985 timeframe.

For projections of other relay satellite capabilities, refer to Subsection 9.1.2. For projected developments in relay satellite subsystems, refer to Subsections 5.1 and 5.2.

5.4 GROUND STATION ANTENNAS

Ground station antennas represent a key link in communication satellite systems, as evidenced by the fact that two of the important measures of communication system performance are directly influenced by the gain of the ground station antenna. These measures of performance are: 1) the EIRP and 2) the figure of merit of a receiving system. The EIRP is expressed in decibels above 1 W (dBW) and is obtained by adding the radiated power in dBW to the gain of the antenna (in decibels above an isotropic source) to which the power is delivered. The figure of merit of a receiving system is the ratio of the antenna gain to the system noise temperature, expressed as dB/K and written as G/T.

Since many other antenna parameters (e.g., bandwidth, sidelobe level, and efficiency) are related to the gain (G) and/or noise temperature (T_a) of the antenna, these two parameters generally describe the performance of a given antenna. The antenna noise temperature, along with the noise temperature of the Low Noise Amplifier (LNA), determine the system noise temperature, which is a good basis for ground station performance.

5.4.1 State of the Art in Ground Station Antennas

The majority of ground station antennas currently in use or planned for the near future use some form of parabolic reflector with diameters ranging from less than 3 m to more than 30 m. Over the past several years, ground station antennas of a given diameter have experienced significant improvement in gain and noise temperature. These improvements have resulted from the development of more efficient feed designs (including corrugated horns and beam waveguides), from the ability to maintain tighter surface tolerances during fabrication and assembly, and from the implementation of surface correcting systems that maintain these tolerances in the field. These improvements in antenna gain and noise temperature, when combined with the increase in available transmitter power (see Subsection 5.1) and the decrease in receiver front-end noise temperature (see Subsection 5.5), have resulted in increased values for ground station EIRP and G/T. Furthermore, the availability of high-gain, low-noise antennas, higher-power transmitters, and ultra-low-temperature receivers has led to cost/configuration tradeoff studies relative to achieving a specified EIRP or G/T. The performance of meaningful tradeoff studies tends to minimize the overall cost (acquisition and ownership) of a ground station designed to achieve a given performance level.

As the more desirable orbital positions of space begin to fill, especially at synchronous altitudes, the demands placed on ground station antennas start to increase. It becomes necessary to maintain lower sidelobe levels and narrower beamwidths. In addition, the shortage of available frequencies creates a need for "frequency reuse" techniques and/or the use of higher frequencies. One common method being used is that of cross-polarization of the signals. By transmitting two signals with orthogonal polarizations (either linear or circular), the satellite can effectively realize a doubling of spectral bandwidth. This technique is currently being used for the 4- to 6-GHz communication band and has been tested extensively at the 11- to 14-GHz Ku-band and at the 30-GHz

satellite band. However, rain causes a major problem when using this technique. The water reduces the isolation between the two orthogonal signals, causing interference between them. Tests indicate that rotating the axis of polarization of the receiving antenna will compensate for part of the problem when using linear polarization, but no solutions have been found for circularly polarized systems. Another frequency reuse system employs spatially isolated antenna beams. In this configuration, a satellite may use the same frequency for two beams, or the same frequency as an adjacent satellite, but with the beams pointed in opposite directions. This method provides a fair amount of isolation between signals, but it introduces more critical requirements for ground station pointing, beamwidth, and sidelobe suppression.

In addition to the standard parabolic reflector ground stations, the technology is available for use by phased-array systems. Several large phased-array radar systems have been established for military operations, and test results indicate that satellite tracking is feasible with a phased-array ground station. A phased-array system allows the use of a large aperture with high gain, since there is no requirement for mechanical steerability. The systems built have exhibited high gain, high power, and low noise response. The current disadvantages of the phased-array system are increased complexity and higher cost.

For general data transmission and reception requirements, it is possible to achieve a given EIRP or G/T with various combinations of antennas and transmitters or LNAs. This makes it difficult to quantize the state of the art of ground station antenna performance. On the other hand, qualitative descriptions can be obtained by plotting representative values of EIRP and G/T as functions of antenna diameter for some existing systems in a given frequency band. Figure 5.4.1-1 presents the EIRP and G/T for typical ground stations operating at S-band. Figure 5.4.1-2 shows the information for C-band stations and Figure 5.4.1-3 covers Ku-band stations. The S- and Ku-band information

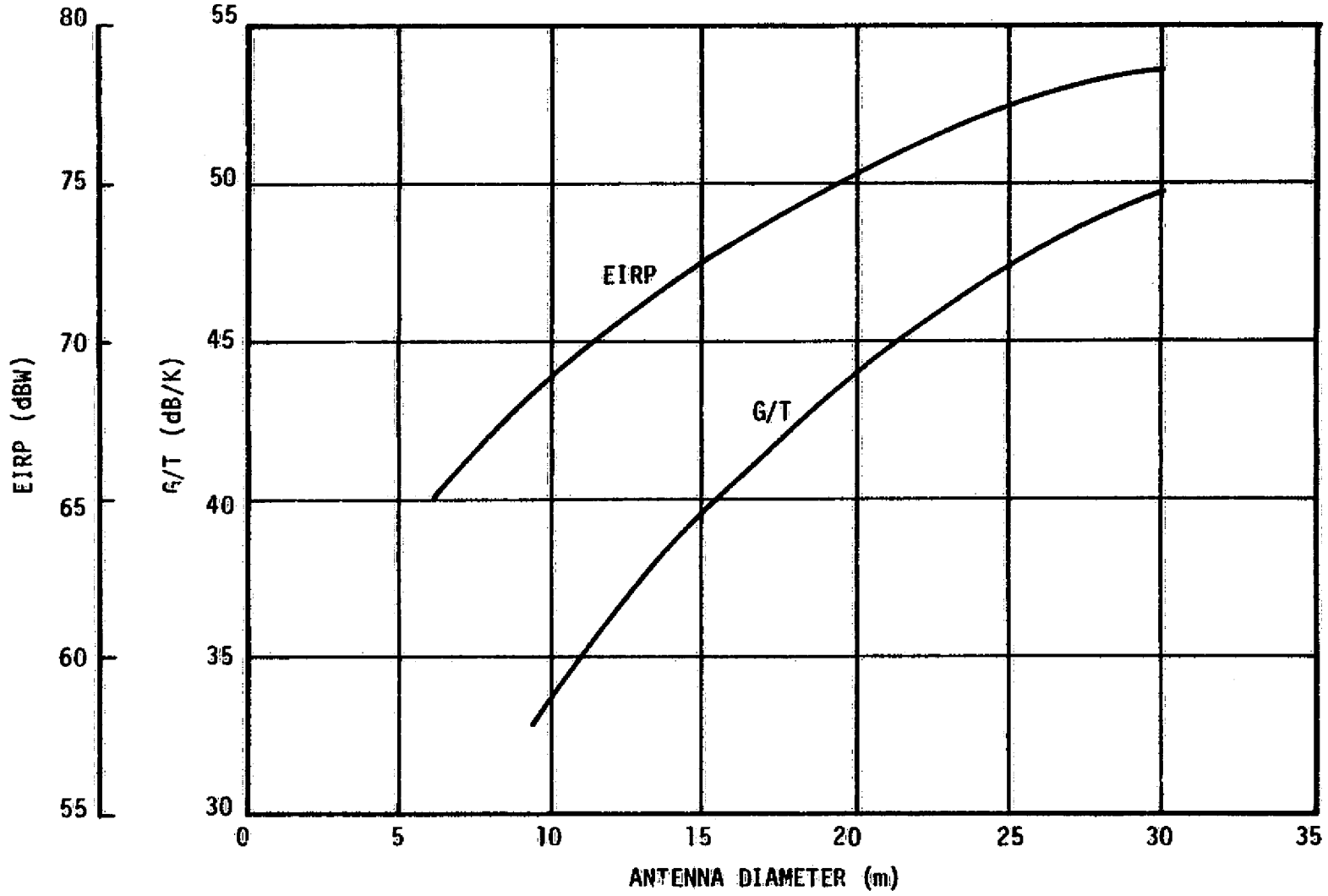


FIGURE 5.4.1-1. REPRESENTATIVE VALUES FOR EIRP AND G/T FOR GROUND STATIONS AT S-BAND (COOLED PARAMETRIC AMPLIFIER)

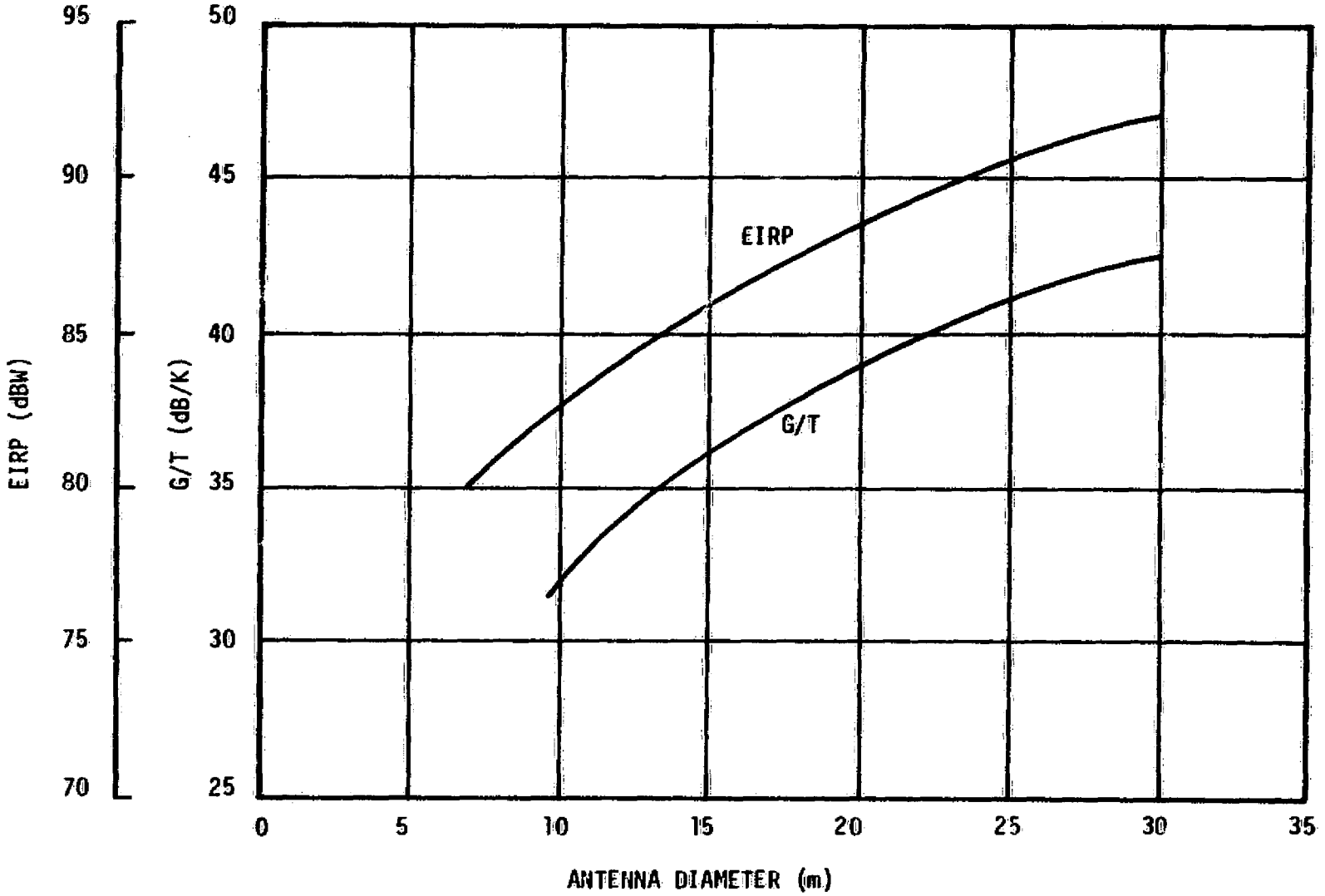


FIGURE 5.4.1-2. REPRESENTATIVE VALUES IN EIRP AND G/T FOR GROUND STATIONS AT C-BAND (UNCOOLED GaAs FET AMPLIFIER)

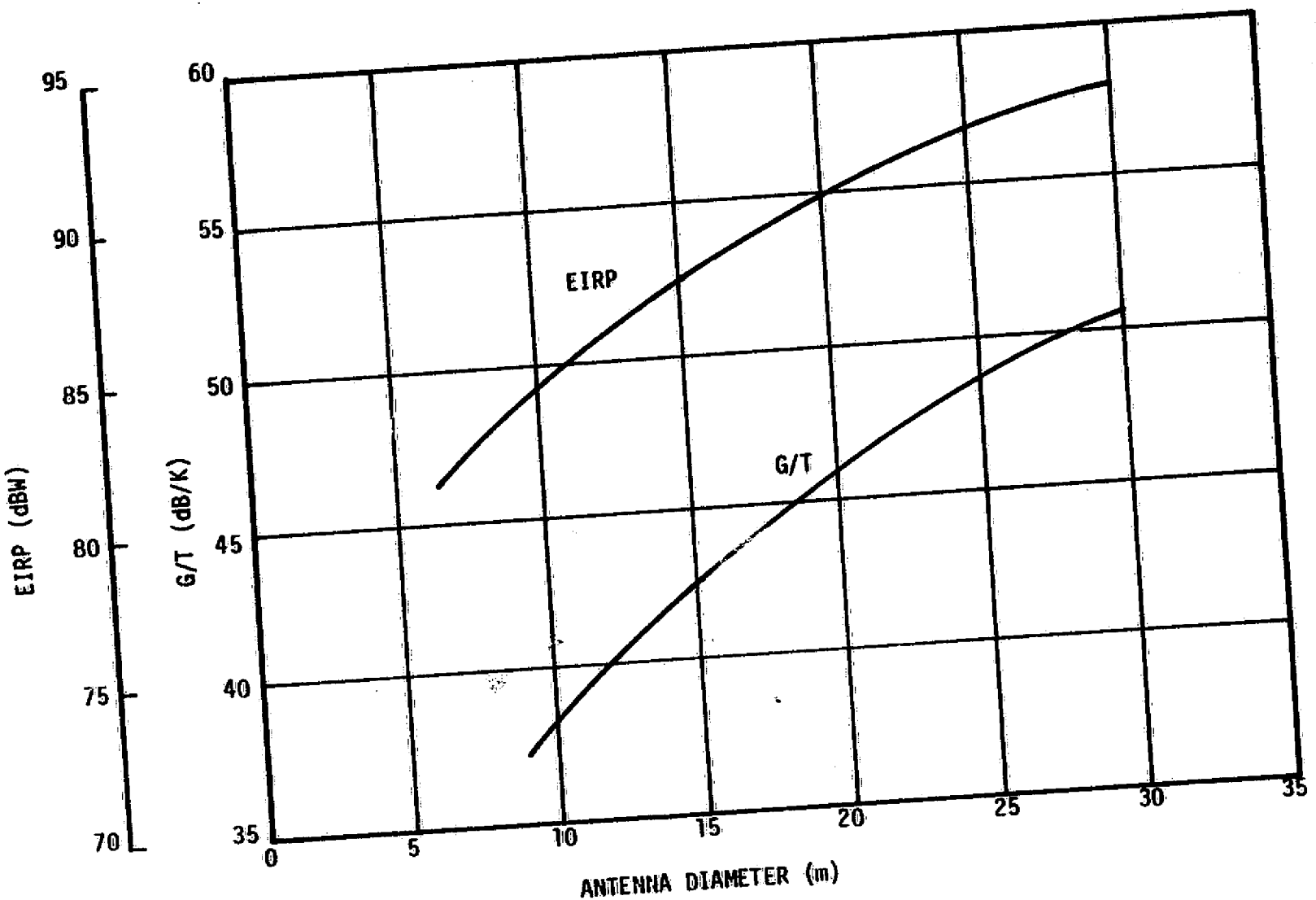


FIGURE 5.4.1-3. REPRESENTATIVE VALUES IN EIRP AND G/T FOR GROUND STATIONS AT Ku-BAND (COOLED PARAMETRIC AMPLIFIER)

is presented for ground stations using cooled parametric amplifiers, whereas the data for C-band is based on communication-type antennas using uncooled GaAs FET amplifiers. Data for systems operating at higher communication bands are currently unavailable.

5.4.2 Trends in Ground Station Antennas

Trends in ground station antenna design will be in the direction that permits increased capacity for existing systems and supports the implementation of additional systems at lower overall costs.

With the continually increasing volume of traffic in the 4- to 6-GHz band, methods of increasing usage of the available spectrum will continue to be developed. More satellites will be implemented with cross-polarization systems, and the ground stations will be designed to take full advantage of this technique. Ground station designs will be required to implement sidelobe suppression and frequency reuse techniques. Ground antennas that are not using state-of-the-art technology will cause severe degradation to the new and somewhat critical frequency reuse designs. Therefore, it is important to maintain an ever-increasing quality of station antennas.

Even with the increased capacity provided by frequency reuse, the 4- to 6-GHz band will lack the capacity and capability to support the increasing demand for additional channels. Both high-speed digital data transmission and broadcast television distribution are shifting from terrestrial to satellite links. As this trend continues, the current band will be pushed to limits, even with frequency conservation and reuse techniques. To avoid this situation, many of the future communication satellites will use the 11- to 14-GHz (SATCOM) band.

The move to these higher frequencies has already started, with several experimental satellites in orbit and several commercial ones in the planning or construction stage. To support the use of the 11- to 14-GHz band, ground station antenna development will cover the entire range of antenna diameters from less than 1 m to greater than 100 m. The larger antennas (greater than 80 m) will be used principally for radio astronomy and extraterrestrial research. For the most part, the high-density, high-speed digital communication traffic will be handled by antennas in the 10- to 20-m range. Smaller antennas in the 3- to 6-m range will be used for television transmission and/or reception

and for low-density FDM and single-channel-per-carrier (SCPC) traffic. Antennas between 0.5 and 2-m in diameter will be used for television reception from satellites. The latter use will depend on achieving very high values of EIRP from the satellites.

Antenna designs based on the parabolic reflector will be the dominant choice for ground station antennas. Some tradeoffs exist in the choice between prime focus (low sidelobes) and Cassegrainian (high gain, high efficiency) feeds, particularly with regard to the smaller antenna sizes. Cassegrainian antennas will become more popular for ground stations as research work continues to reduce sidelobes. Offset feed and multiple-beam designs will become popular. In addition, the need to reduce outages (caused by Sun transit, satellite failure, etc.) will require the development of steerable-beam feeds for use with the larger reflectors. At higher frequencies, beam waveguides will become more common, since they reduce the structural requirements on the antenna and its mount and offer the potential for low loss.

Cost reductions will be realized through the use of more reliable equipment, including antenna steering and control systems, small fixed-position receiving antennas with geosynchronous satellites, and a trend toward standardized rather than one-of-a-kind ground stations.

5.4.3 Projected Developments in Ground Station Antennas

Future developments in ground station antennas will be related to the continued increase in frequency of the satellite communication systems. An increase to the 20- to 30-GHz band, where the International Telecommunications Union (ITU) has allocated contiguous bandwidths of up to 2,500 MHz, is inevitable. Ground station antennas at these frequencies will use the same design techniques that have been developed for the lower bands. The required tolerances on these antennas will increase, however, with surface and pointing tolerances being major design drivers.

At the 4- to 6-GHz and 11- to 14-GHz bands, the antennas will employ frequency reuse techniques. Pointing tolerances will be stricter and sidelobes will be lower. The G/T for ground stations will be exceptionally good as lower noise receivers are built. Many of the ground stations will use low-noise, uncooled GaAs FETs in their associated amplifiers. Multiple-beam and steerable-beam antennas will come into general usage. Most of these designs will continue to incorporate the use of parabolic reflectors. Phased arrays will find only limited use because of higher cost, restricted field of view ($\pm$60 deg), and the lack of a need for continuously variable tracking in most applications.

5.5 MICROWAVE RECEIVERS

As the trends in space communications migrate to higher data rates over greater transmission distances, the need for improved receiver performance assumes increasing importance. Frequently, the performance improvements produce conflicting demands on receiver design, such as improved noise figure with increasing bandwidths. Such demands obviously require technology advances in receiver front ends, higher receiver frequencies, frequency selection techniques, and improved oscillator stability, as well as the usual physical considerations.

The microwave receivers discussed herein include that part of the overall receiving system between the antenna and the demodulator. Thus, the receiver defined in this subsection provides gain, frequency selection, frequency conversion, and bandwidth limiting. Demodulation of the signal is discussed in Subsection 5.6.

5.5.1 State of the Art in Microwave Receivers

Data receivers are designed to operate within a multiuser frequency spectrum, and the permissible bandwidths are standardized by Inter-Range Instrumentation Group (IRIG) regulations. The restrictions imposed by these regulations generally preclude bandwidth from consideration as a technology item (use of available bandwidth is a function of the modulation and coding scheme, and is discussed in Subsection 5.6). The primary items of concern from a technology viewpoint then become receiver frequency, receiver noise figure as a function of frequency, local oscillator stability, frequency selection techniques, and physical considerations, including size, weight, power consumption, system configuration, and environmental limitations. These receiver parameters are shown in Figure 5.5.1-1 for clarification.

The availability of efficient, reliable hardware at frequencies above 10 GHz is of primary concern to designers of future systems because of bandwidth crowding at the lower frequencies and because of the wider bandwidths associated with the higher frequencies. Significant improvements in both commercially available and developmental front end devices have occurred within the past 2 years, particularly in the area of GaAs FETs, with some improvements in extending the upper frequency of cryocooled parametric amplifiers to include the 35- to 36-GHz band. Figure 5.5.1-2 presents curves of noise temperature as a function of frequency for various devices commercially available in 1977. Typical values of receiver noise temperature for several types of front ends are given in Table 5.5.1-1. The values presented in this table represent commercially available hardware.

Recent developments in low-noise amplifier (LNA) designs include uncooled parametric amplifiers having noise temperatures of 40 to 100 K and operating in the 2-to-15-GHz region. High-performance, ultra-low-noise parametric amplifiers have achieved noise temperatures of 25 to 70 K in the 10- to 40-GHz region. In addition, low-noise commercially available GaAs FETs have achieved noise temperatures as low as 120 K at 2 GHz, with several developmental units approaching 100 K. A listing of commercially available GaAs FETs, along with their specifications, is presented in Table A-6 of Appendix A. This table indicates a lack of commercially available GaAs FETs at frequencies above 12 GHz.

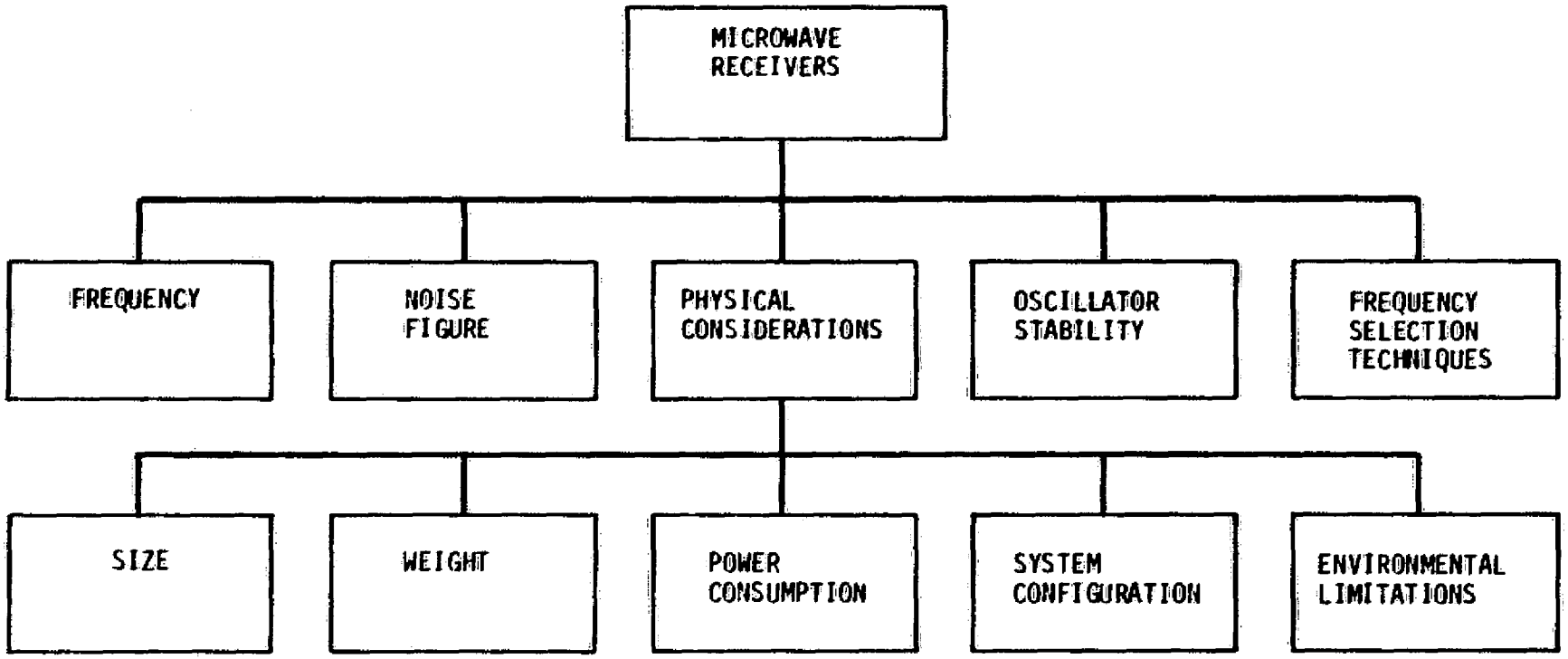


FIGURE 5.5.1-1. MICROWAVE RECEIVER PARAMETERS

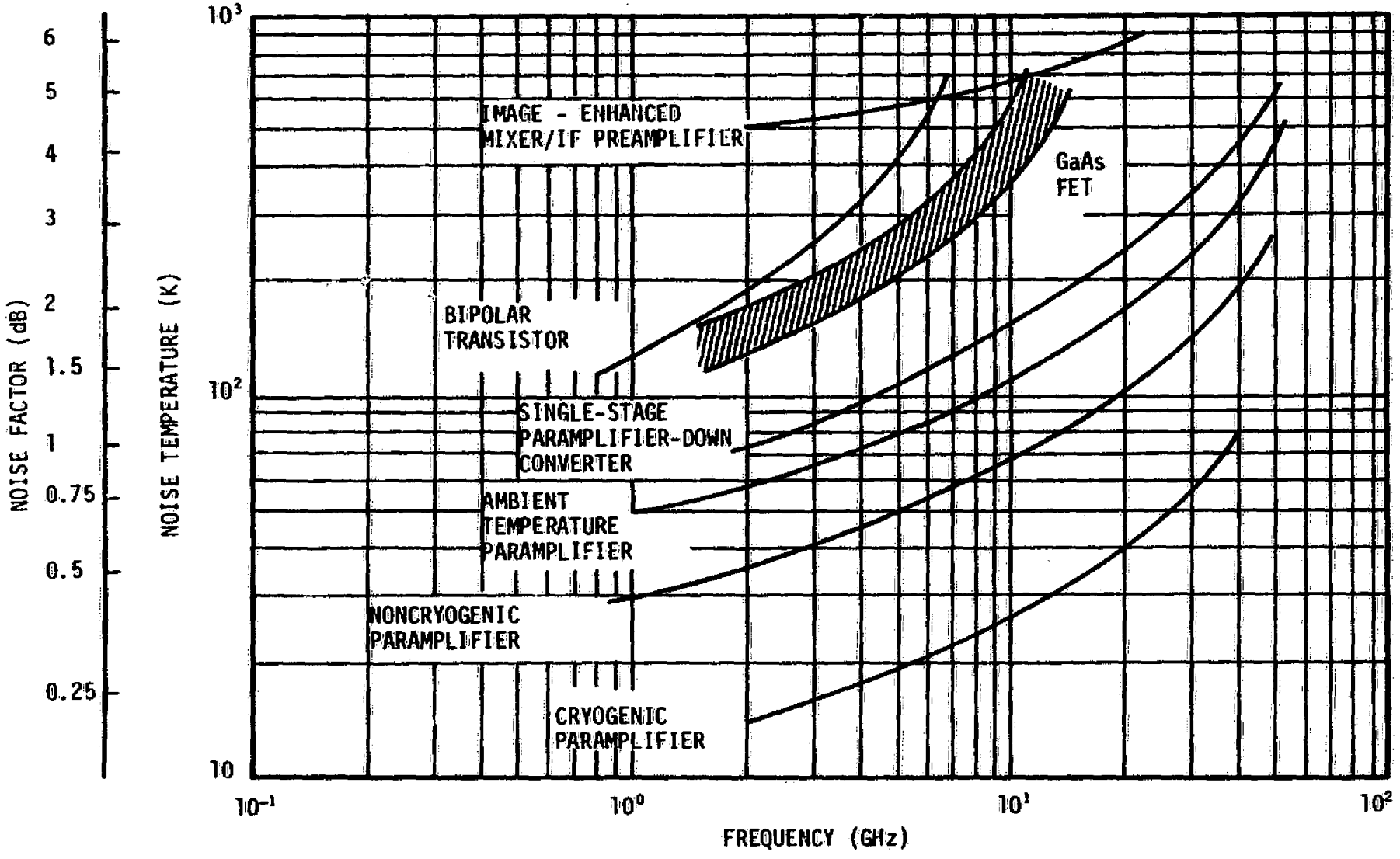


FIGURE 5.5.1-2. STATE-OF-THE-ART RECEIVER FRONT END NOISE TEMPERATURE AS A FUNCTION OF FREQUENCY

TABLE 5.5.1-1. 1977 STATE OF THE ART IN RECEIVER NOISE TEMPERATURE

FRONT END DEVICE	CHARACTERISTIC	
	NOISE TEMPERATURE (K)	FREQUENCY (GHz)
Bipolar Transistor	125 320 500	1 4 6
GaAs FET	120 200 350	2 5 10
Paramplifier (Electrically Cooled)	40 65 170	4 10 36
Paramplifier (Cryo-Cooled)	17 25 65	4 10 36

Several recently developed low-noise amplifiers are listed in Table 5.5.1-2. The table includes a space-qualified, low-noise, S-band parametric amplifier that has 18-dB gain and a noise temperature below 50 K. The unit weighs 21 oz and consumes 8 W of power. In the 4-GHz satellite band, several "small terminal" type amplifiers are available in the 75- to 85-K range. For larger 4-GHz terminals, a parametric amplifier designed for "hub-mounting" can give noise figures of 45 K. At Ku-band, recent developments have produced a two-stage, 26-dB gain amplifier for operation in the 14.7- to 15.2-GHz band. The noise temperature of the unit is 82 K and the 1-dB bandwidth is 650 MHz. Finally, at Ka-band, there is an electronically tunable paramplifier down converter operating from 36.5 to 38.5 GHz with a noise temperature of 300 K and an instantaneous bandwidth of 150 MHz.

In general, data tend to indicate that receiver front ends at the higher frequencies (above 40 GHz) do not use RF amplifiers. Instead, the signal is fed directly from the preselector to the mixer. As low-noise amplifiers become available at these frequencies, the trend will be to amplify before mixing, as is now done at the lower frequencies.

TABLE 5.5.1-2. STATE OF THE ART FOR UNCOOLED LOW-NOISE AMPLIFIERS

DEVICE	FREQUENCY (GHz)	GAIN (dB)	NOISE TEMPERATURE (K)
Paramplifier	2.2 to 2.3	18	50
GaAs FET	3.7 to 4.2	TBD	75 to 85
Paramplifier	3.7 to 4.2	16	45
Paramplifier	14.7 to 15.2	26	82
Paramplifier Down Converter	36.5 to 38.5	TBD	300

In the area of local oscillators, several options are available to the user. Frequency synthesizers offer a convenient but expensive method of implementing variable frequency receivers. By programming the synthesizer, it is possible to tune the receiver to any number of desired frequencies. Frequency synthesizers are available for most of the common microwave communication bands, including the 18- to 40-GHz frequency range. Several available units offer good stability, low noise, and broad frequency coverage. Whether the oscillator frequency is derived from a synthesizer or on a fixed-frequency basis, it is necessary to have a stable local oscillator. The state of the art in local oscillator stability is presented in Table 5.5.1-3 for the three more commonly used sources. Of these three sources, the cesium source is a primary standard, meaning that it does not require reference calibration. The rubidium and crystal oscillator sources are secondary standards and must be calibrated periodically. Because of cost limitations, quartz crystals are the most commonly used sources. Highly accurate quartz oscillators with stabilities of $\pm 5 \times 10^{-10}$/day are available for medium-tolerance frequency control, whereas less costly models are available with stabilities on the order of $\pm 1 \times 10^{-6}$/day. The use of rubidium or cesium sources is generally limited to applications requiring extreme accuracy (e.g., tracking receivers).

TABLE 5.5.1-3. COMMERCIALLY AVAILABLE OSCILLATOR CHARACTERISTICS

CHARACTERISTIC	CESIUM	RUBIDIUM	QUARTZ
Accuracy.	$\pm 7 \times 10^{-12}$	N/A	$\pm 2 \times 10^{-6}$
Long-Term Stability	$\pm 3 \times 10^{-12}$/life	$\pm 1 \times 10^{-11}$/month	$\pm 5 \times 10^{-10}$/day
Short-Term (1 sec) Stability	5.0×10^{-12}	5×10^{-12}	1×10^{-11}
Relative Cost ($)	23K	9K	600
Relative Size (in.)	9 by 17 by 16	5 by 17 by 16	2.8 by 2 by 2.4
Relative Weight (lb)	70	34	0.7

5.5.2 Trends in Microwave Receivers

Continuing pressure to improve performance has motivated manufacturers to find more effective methods of implementation for receiver hardware. This has resulted in placing more functions per module and reducing the life-cycle costs of microwave systems. The combining of a number of active and passive components that would normally be packaged separately has led to the "super component". An example of a super component at Ku-band is the integration of six component functions into one box. An input switching matrix, limiter, voltage variable attenuator, coupler, detector, and output switching matrix can be incorporated into one package about the size of a discrete-voltage variable attenuator. The resulting circuit not only saves weight and size but reduces the external interconnections and their associated losses. Tolerances are more strictly controlled, and the end product is far superior to a discrete design. Costs of the super components remain about the same as for discrete designs.

Increased use of Large-Scale Integrated (LSI) circuits will allow increased capability, improved reliability, reduced maintenance cost, lower power consumption, and a reduction in size. Items such as Charge-Coupled Devices (CCDs) and Surface Acoustic Wave (SAW) filters have the potential to provide further reductions in receiver size, while providing very stable, maintenance-free bandwidth filtering. Technology is moving in the direction of developing an entire receiver on one substrate. However, unless there is a greater standardization of receiver IF bandwidths, the use of SAW devices and integrated IF modules will not be cost-effective. (A system such as commercial television has a broad, standardized market that will enable SAW devices to become common.)

Further development of gallium arsenide (Ga) and indium phosphide (InP) FETs will allow the replacement of low-noise TWTs at frequencies below 18 GHz. This will raise receiver efficiency, lower receiver power, and allow more reliable operation. As microelectronics continue to develop, the trend will be toward "smart

receivers". Under microprocessor control, the receiver would be capable of precise frequency locking and could exhibit a variable bandwidth IF optimized to signal conditions.

With the use of higher receiver frequencies, new techniques will be developed in the millimeter and submillimeter regions. Currently available mixer diodes are capable of detecting signals at terahertz frequencies, but conventional receiver designs limit the frequency to 100 GHz. Development of "quasioptics" and new chip mounting techniques will allow receivers to have frequency capability approaching the infared region (laboratory models reaching 670 GHz have already been demonstrated).

Trends in the noise temperatures of several devices are presented in Figures 5.5.2-1 through 5.5.2-4. These predicted trends are based on the recently developed low-noise amplifier designs, as well as work currently being done at the laboratory development level. The trends show noise temperature being reduced for all devices at all frequencies. The greatest reductions are for GaAs FETs at all frequencies between 2 and 18 GHz, and for the paramplifier devices at frequencies above 30 GHz.

With the incorporation of super components and LSI, receiver size will decrease greatly during the 1980 to 1985 timeframe, being reduced to half-size by 1985. In addition, the use of uncooled solid-state amplifiers will reduce the power consumption at the same time that size is reduced. The typical receiver power consumption by 1985 will be somewhat below current power requirements, even though the receiver capability will be increased. Costs for receivers will not vary greatly, with the lower costs of integrated functions being offset by the higher cost of design development.

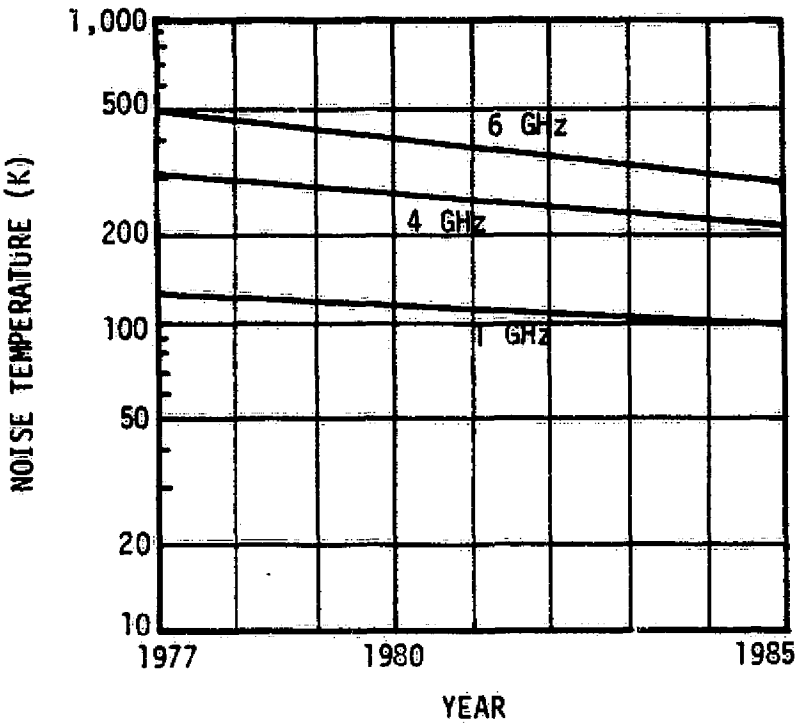


FIGURE 5.5.2-1. PROJECTED TRENDS IN BIPOLAR NOISE TEMPERATURE

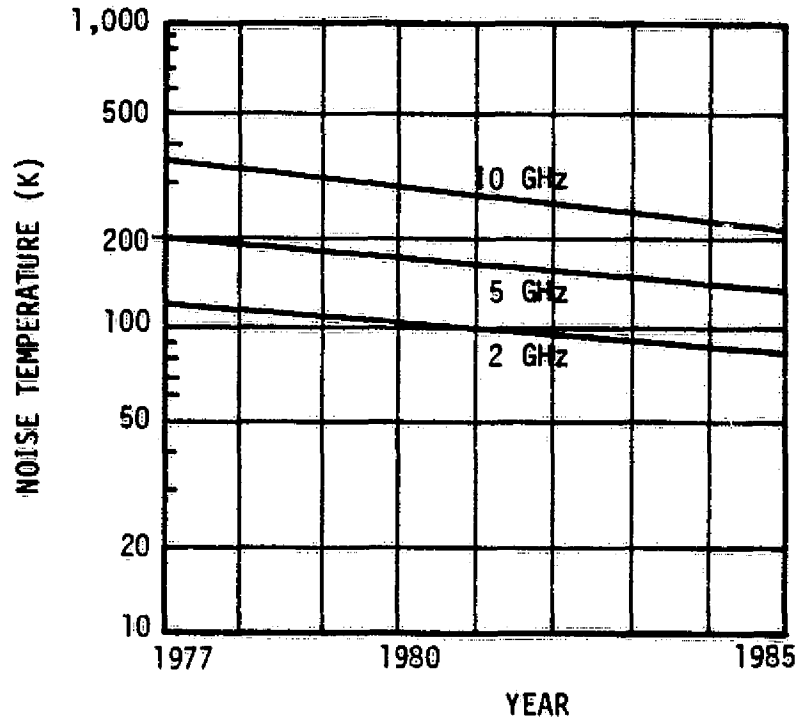


FIGURE 5.5.2-2. PROJECTED TRENDS IN GaAs FET NOISE TEMPERATURE

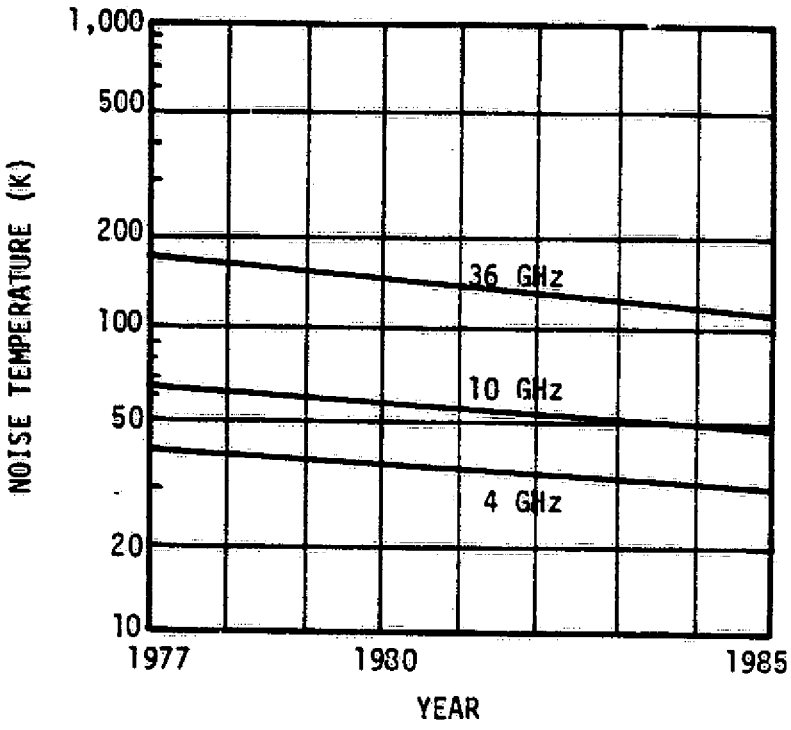


FIGURE 5.5.2-3. PROJECTED TRENDS IN ELECTRICALLY COOLED PARAMPLIFIER NOISE TEMPERATURE

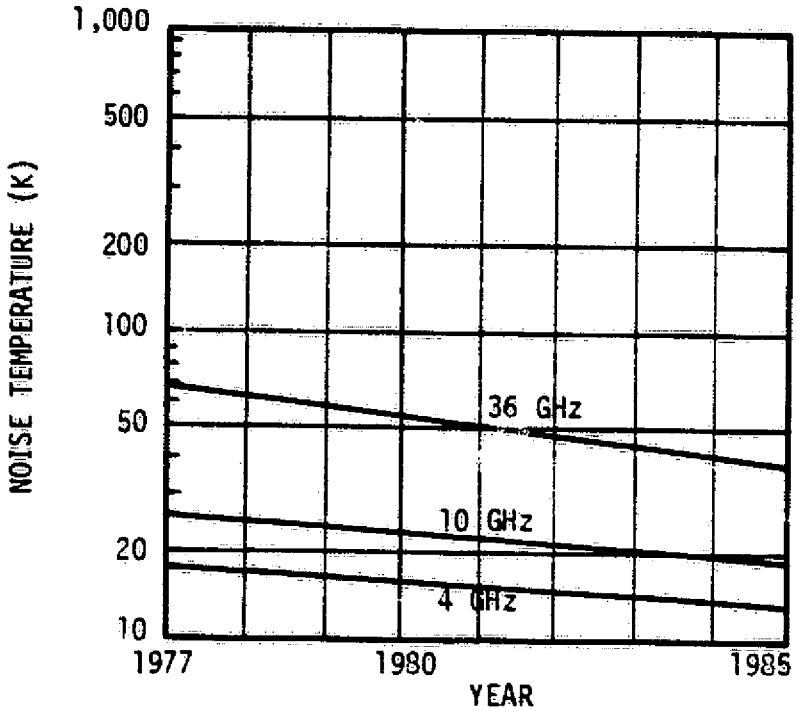


FIGURE 5.5.2-4. PROJECTED TRENDS IN CRYO-COOLED PARAMPLIFIER NOISE TEMPERATURE

5.5.3 Projected Developments in Microwave Receivers

The receivers of the 1980-1985 timeframe will be capable of providing lower noise operation while handling increased data rates. Projections of receiver noise temperature for 1980 and 1985 are given in Tables 5.5.3-1 and 5.5.3-2. These projections are based on recent trends in low-noise amplifiers and expected developments as a result of current laboratory research.

The use of LSI, SAW bandwidth filters, and microprocessor controllers will allow increased capability by optimizing the receiver to the signal. Super components will reduce receiver size and increase system reliability and performance. Reliability and performance will also be increased by the replacement of TWTs by solid-state devices at frequencies of Ku-band and below. Many of the environmental restrictions of receivers will be reduced as the cryogenically cooled and electrically cooled parametric amplifiers are replaced by ambient-temperature parametric amplifiers and GaAs FET amplifiers. The frequency ranges of parametric amplifiers and FETs will increase, allowing pre-mixer amplification at frequencies well above Ka-band. For frequencies in the millimeter and submillimeter regions, quasioptics and new chip mounting techniques will allow reliable receiver operation with very high bandwidths (tens of gigahertz) available.

TABLE 5.5.3-1. 1980 PROJECTIONS OF RECEIVER NOISE TEMPERATURE

FRONT END DEVICE	CHARACTERISTIC	
	NOISE TEMPERATURE (K)	FREQUENCY (GHz)
Bipolar Transistor	115 270 400	1 4 6
GaAs FET	105 170 290	2 5 10
Paramplifier (Electrically Cooled)	36 57 145	4 10 36
Paramplifier (Cryo-Cooled)	15 22 55	4 10 36

TABLE 5.5.3-2. 1985 PROJECTIONS OF RECEIVER NOISE TEMPERATURE

FRONT END DEVICE	CHARACTERISTIC	
	NOISE TEMPERATURE (K)	FREQUENCY (GHz)
Bipolar Transistor	100 220 300	1 4 6
GaAs FET	85 135 210	2 5 10
Paramplifier (Electrically Cooled)	30 46 110	4 10 36
Paramplifier (Cryo-Cooled)	13 18 40	4 10 36

5.6 MODULATION, DEMODULATION, AND CODING

Modulation, demodulation, coding, and detection are the major areas of signal design and processing for data communication systems. The first three areas are discussed in this subsection. Detection has arbitrarily been relegated to Subsection 6.1.

As data rates increase, requirements for acceptable bit error performance do not lessen, and in fact frequently increase; yet the bandwidth crowding and lower bit dwell times for high-rate data tend to cause even more errors. In most cases the increased data must be accommodated on a channel of fixed bandwidth; therefore, efficient use of assigned channel space becomes a critical factor. As these demands on the data systems increase, the modulation/demodulation and coding/decoding processes frequently become the only media for further increases and thus these processes take on increasing importance with each demand for more efficient channel utilization. A number of techniques to reduce the required bandwidth-to-bit ratio have been developed over the past 20 to 25 years, but many of these techniques have not been feasible to implement until recently, when microminiaturization of electronic devices became feasible. This subsection addresses some of the techniques that are currently feasible and projects the trends and developments that are expected as further improvements in microelectronics take place.

5.6.1 State of the Art in Modulation, Demodulation, and Coding Technology

The processes of coding/decoding, modulation/demodulation, and bandwidth compression are included in the category of signal design techniques and are discussed herein as methods rather than hardware.

5.6.1.1 Modulation - Modulation characteristics of interest are shown in Figure 5.6.1.1-1. The two major classes of modulation are angle and amplitude modulation. Angle modulation is further broken into Phase Modulation (PM) and Frequency Modulation (FM). The factors of bandwidth, linearity, and implementation cost apply to all forms of modulation. The use of Amplitude Modulation (AM) for data transmission is generally restricted to the optical band (laser) in today's systems because of its simplicity of implementation with optical devices. A few terrestrial data systems use a combination of AM and PM, but these techniques have not been applied to spacecraft data links. The use of FM for data transmission is generally restricted to Frequency Division Multiplexing (FDM)/FM systems where data rates are relatively low (less than 50 Kbits/sec). The most common form of communication data modulation is PM, finding use from the lowest to the highest data rates.

The more common type of PM is Binary-Phase-Shift-Keyed (BPSK) modulation, a technique in which the data modulates an RF carrier to 0- or 180-deg phases [also called Phase Reversal Keying (PRK)]. The use of BPSK is well-established, both for space data transfer and terrestrial systems. To transfer more data in a limited amount of bandwidth, it is possible to introduce more phase angles into the modulation system. Quadriphase-Shift-Keying (QPSK) involves one binary bit stream modulating a carrier between 0 and 180 deg while another bit stream of the same rate is modulating the carrier between 90 and 270 deg. This effectively allows the transfer of twice the data rate within the given amount of RF bandwidth required for BPSK.

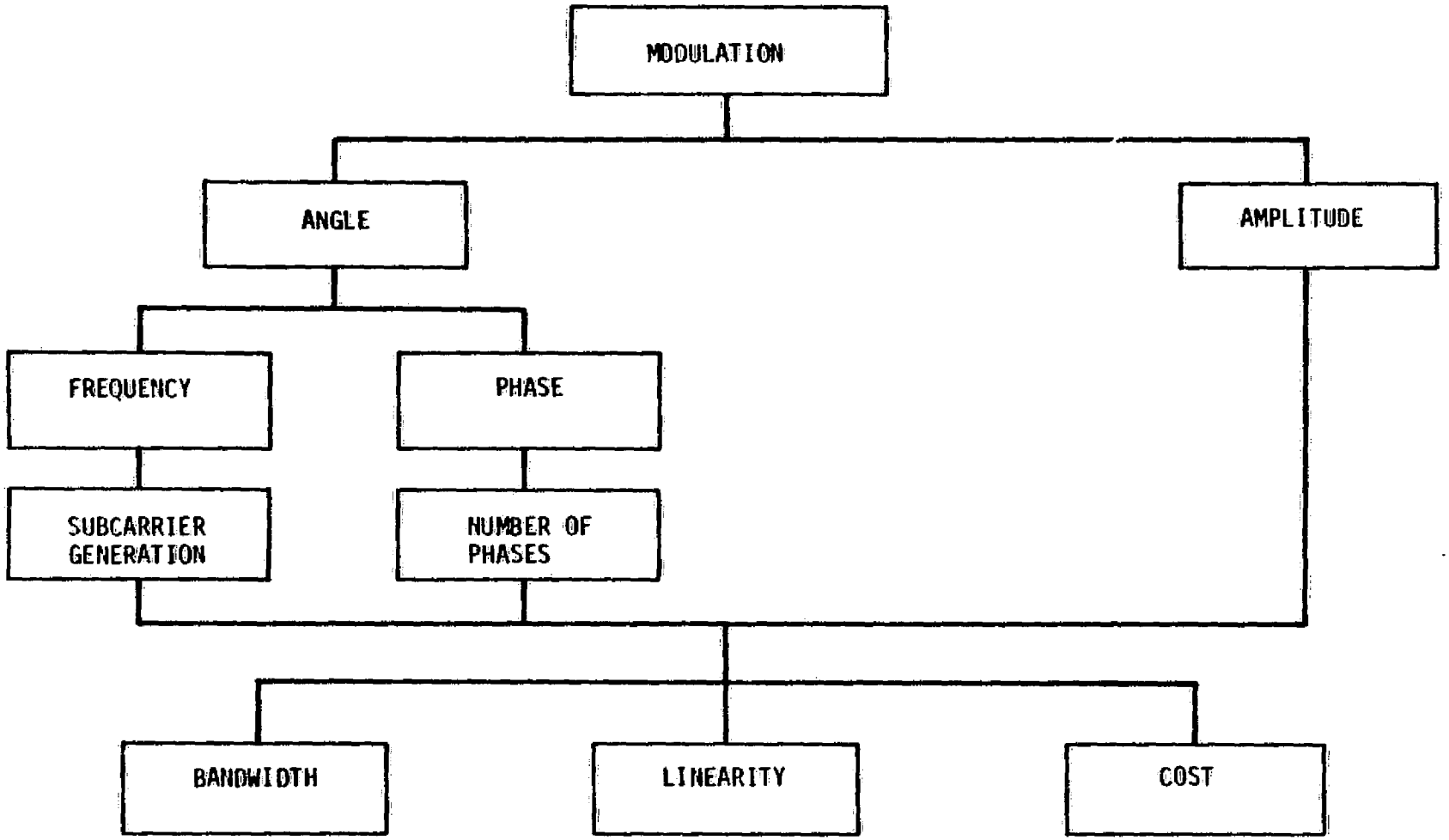


FIGURE 5.6.1.1-1. MODULATION CHARACTERISTICS

The need for transferring more data through a finite bandwidth has resulted in another measure of technology, i.e., bits-per-second-per-hertz (bits/sec/Hz). Ordinary BPSK modulation results in 0.5 bits/sec/Hz (two cycles of bandwidth for each bit of data). QPSK modulation doubles the rate, allowing 1 bits/sec/Hz. Systems using 8- and 16-level PM schemes increase this number, as shown in Table 5.6.1.1-1. Table 5.6.1.1-2 lists the bandwidth efficiency as a function of modulation technique for a number of operational and experimental systems currently in use (Ref. 5-2).

As the required bandwidth decreases, the Bit Energy-to-Noise Ratio (E_b/N_o) must increase to maintain the same Bit Error Rate (BER). For example, for a BER of 10^{-6}, the required E_b/N_o goes from 10.6 dB for BPSK (Table 5.6.1.1-1) to 18.2 dB for 16-level PM. This places a restriction on th system by requiring an improvement in signal-to-noise techniques. Additionally, in a multiuser environment, the multiphase modulation techniques are more susceptible to errors, since any small shift of carrier phase will be detected as data, even though the shift results from an interfering signal. As a result of these restrictions, use of multiple-phase modulation will likely be limited to QPSK for space applications.

QPSK can be accomplished by using two bi-phase modulators. For low-power and low-to-intermediate-frequency systems (at X-band and below), a bit stream modulates each BPSK unit, with the resulting outputs combined in parallel. This approach has been used for data rates of up to 1 Gbits/sec. Another approach is to BPSK modulate the carrier with one bit stream and to then BPSK modulate the resulting carrier with a second bit stream. This series QPSK modulation technique adapts readily to waveguide circuits and can be implemented with PIN diodes, Schottky diodes, or FETs. In addition, these methods can be implemented on chips, with current capabilities being the modulation of a 5-GHz carrier using a 1-Gbits/sec data stream.

NASA's Goddard Space Flight Center has recently completed a 5-year study of satellite-to-satellite laser communication systems. Data rates of up to 300 Mbits/sec have been achieved using a CO_2 laser with a 0.95-W

TABLE 5.6.1.1-1. CHARACTERISTICS OF M-ARY-PHASE-SHIFT-KEYED MODULATION SCHEMES

MODULATION TYPE	EFFICIENCY* (bits/sec/Hz)	BANDWIDTH PER BIT	REQUIRED E_b/N_0 FOR BER = 10^{-6} (dB)
BPSK	0.5	2.0	10.6
QPSK	1.0	1.0	11.4
8-PSK	2.0	0.5	12.7
16-PSK	4.0	0.25	18.2

*For typical ground-based transmission systems.

TABLE 5.6.1.1-2. BANDWIDTH EFFICIENCY AS A FUNCTION OF MODULATION TECHNIQUE (REF. 5-2)

LEVEL	TYPE MODULATION	bits/sec/Hz	WHERE USED
2	2 ϕ PSK (using M)	0.50	Canada Thin Route Norway Marisat TTY
4	QPSK QPSK QPSK SQPSK QSPSK FFSK	0.94 1.12 1.53 1.30 1.80 2.20 (2.5 capability)	SBS SPADE (Intelsat) Telsat Bell System T3 transmission via 11-GHz radio Experimental (F. Chethik) CTS Canada Experiment (11/14-GHz through 85-MHz BW)
6	QPRK	2.25	Microwave Associates 11-GHz terrestrial radio
7	Zero-memory Nyquist correlative coding: modulating a PCM-FM system	4.00	Lenkurt system for transmitting 6.312 Mbits/sec T2 over 3-MHz terrestrial radio channel
8	8 ϕ PSK 8 ϕ PSK	1.81 2.20	TDMS through Intelsat IV by Fujitsu Ltd. NEC terrestrial radio for 11 GHz
16	16 ϕ PSK 16 APSK*	4.00 4.00	Comsat Laboratories/ Japan NTT/ECL Japan NTT/ECL+
32	4-bit QASK**	4.00	JPL for Space Lab++
49	Hexagonal type signal format using superimposed modulation	4.00	Japan NTT/ECL+ (experimental)

* Combined amplitude and phase shift keying
**Quadrature amplitude shift keying
+ K. Myauchi, IEEE Trans Comm., February 1976
++J. Smith, AIAA paper 76-230

output. Modulation is achieved using a cadmium telluride rod to change the polarization of the laser output (polarization modulation). By using linear polarization for the system, the change in polarization caused by modulation will either increase or decrease the received signal, resulting in AM. However, if the light output is circularly polarized, the change in polarization (due to the modulation) will be detected as a change in frequency (FM).

5.6.1.2 Demodulation - Demodulation, as covered in this subsection, provides the conversion of the IF output of the receiver into a baseband signal. Demodulation has the same characteristics as were illustrated for modulation in Figure 5.6.1.1-1.

The major type of demodulation used in the past for missile telemetry systems was FM. It is still used for FDM/FM signals of (now) relatively narrow bandwidth. The FDM baseband as defined in IRIG-106-75, "Telemetry Standards" (Ref. 5-3), is limited to about 200 kHz. This FDM system allows the user a maximum of 21 analog data channels with frequency responses from 6 to 2,475 Hz. PM demodulation has been used mainly for PCM/PM signals since the trend to coherent digital transmission of signals has become predominant. AM for space data transmission is currently limited to the optical band. (A combination AM/PM is used for some high-data-rate terrestrial cable systems.)

BPSK modulation is very common for low to intermediate binary data rates. BPSK is a specific case of PCM/PM where the binary data modulates the RF carrier to 0- or 180-deg phases. Demodulation of BPSK requires a carrier reconstruction (i.e., a carrier without PM). This is accomplished by multiplying the received signal by two and phase-locking the local oscillator to it. The doubled signal modulation is modulus-360 and thus the doubled carrier appears to have no modulation. QPSK modulation is also used where the carrier is modulated by one binary bit stream between 0 and 180 deg while also being modulated by a second binary bit stream of the same bit rate between 90 and 270 deg. Demodulating QPSK requires that the carrier be reconstructed in a manner similar to BPSK. This is accomplished by multiplying the received intermediate

frequency signal by four to phase-lock a local oscillator to it. The X4 multipliers use step-recovery diodes, tapped delay-line frequency multipliers, or FETs. Demodulation of the higher phase systems (8- and 16-ϕ PSK) is accomplished in a manner similar to BPSK and QPSK systems; however, these higher phase systems are currently used only for terrestrial data transfer.

5.6.1.3 Coding - The encoding of digital data began with the use of binary bit codes (e.g., NRZ, RZ, bi-phase) to vary the characteristics of the spectra produced. Block codes (e.g., adding parity bits to a word or frame) were introduced as a means of detecting bit errors and thus increasing the reliability of the data. Continuous codes [e.g., convolutional and Pseudo-Random Noise (PN)] were later used for improving data reliability, spectrum spreading, signal identification, and measuring satellite range and range rate.

Certain communication links under development, such as the TDRSS, require the use of concatenated (multiple) codes. In concatenated codes, the data are encoded using one code and the resultant bit stream is again encoded to produce a layered effect. In the case of the TDRSS, a 2^{18} bit PN code is used for spectrum spreading, range and range-rate information, and user identification. This PN code is then convolutionally encoded to improve the data reliability for transmission over the long distances required. In decoding the data, the convolutional code (last code applied) is first removed, leaving a high-quality PN coded signal. The PN code is then decoded and the transmitted information can be obtained from the resulting data.

In the past, power limitations on the transmitted signal resulted in the additive noise being the predominant cause of random errors. This limitation on power no longer exists for many applications, and the once-frequent random errors have become less significant than burst errors (loss of multiple bits in a group) and slippage errors (addition or deletion of a bit). The convolutional codes currently used are capable of correcting random errors, but they must be interleaved to correct long burst errors and are of little value for correcting slippage errors. Block codes such as the BCH code and the Reed-Solomon code are capable of correcting all three types of errors. The use of these codes, however, is declining, since speed constraints limit their usefulness at high data rates.

5.6.2 Trends and Projected Developments in Modulation, Demodulation, and Coding

The trend over the years has been to transfer information at ever-increasing rates. In the 1960's and even early 1970's, FDM of analog signals was quite common. RF bandwidths of 400 kHz or less were adequate in most cases. More and more, analog signals are being converted to digital signals before transmission. Once converted to digital signals, multiple-phase modulation schemes are being used extensively to increase the bits/sec/Hz that are transmitted over available bandwidths. The typical data rates for ground systems are expected to increase to around 1 Gbits/sec by 1985. The data rate for civilian space systems is expected (with rare exceptions) to remain limited to 300 Mbits/sec, first because of bandwidth limitations of the TDRSS, and second because onboard processing (data set selection, compression, adaptive sensors, etc.) can be implemented to reduce the data output of future high-rate sensors.

Modulation schemes for data transmission through 1985 will tend to divide into two types: 1) low-data-rate, narrowband systems and 2) high-data-rate, wideband systems. The latter systems will use M-ary PM schemes for frequencies below the optical band. The optical band will continue to use AM. Data rates for both the optical and RF bands will be in the range of 200 to 1,000 Mbits/sec for terrestrial use and 300 Mbits/sec for space use.

Demodulation will be performed on higher data rates at greater IF frequencies and wider IF bandwidths. Space data demodulation for the 1980-1985 timeframe will be mainly of the PM type, with BPSK and QPSK being most common. Implementation of 8- and 16-ϕ PSK systems will be unlikely before 1985. The use of FM demodulation for space data transfer will see decreasing use, with the main application being TDM/FM of ground data through domestic communication satellites. The use of AM demodulation for space data transfer will be limited to the optical band.

The demodulation schemes of the 1980's will be implemented with SAW devices for fixed-tuned bandpass filtering and CCDs for transversal filtering. Additionally, demodulation will be performed with large-scale integrated devices, allowing complex systems to be built with a reduction in size and power consumption. This is especially true for the high-rate, M-ary PM techniques.

In the area of coding, the trend is toward convolutional codes for error correction and concatenated codes for multipurpose applications and extremely high-quality data. Both the Space Shuttle and the TDRSS utilize convolutional codes for error detection and correction. Compatibility with TDRSS and Shuttle will necessitate the use of these continuous codes in future spacecraft systems. Since coding schemes must be responsive to the error phenomenon, future systems that use higher frequencies will not necessarily use the same techniques as those at lower frequencies. The higher frequencies are more susceptible to fading and atmospheric disturbances, and errors encountered at these frequencies will probably result from both bursts and long fades. To reduce errors at these frequencies, it may be necessary to employ concatenated codes, with one code correcting the short burst errors and another correcting errors caused by weak signals during the long fades. Finally, the use of "Packet Telemetry" will allow the repetition of any group of data that cannot be decoded without errors.

REFERENCES - SECTION 5

5-1 "Tracking and Data Relay Satellite System (TRDSS) Users' Guide, STDN 101.2", Goddard Space Flight Center, Maryland, January 1978

5-2 C. L. Cuccia, "New PCM Techniques Stress $ and Spectrum Conservation, Part II", Microwave Systems News, p. 38, February 1977

5-3 "Telemetry Standards", IRIG-106-75

5-4 "FET Scoreboard", Microwave Systems News, pp. 70-71, February 1977

BIBLIOGRAPHY - SECTION 5

E. D. Cohen, "TRAPATTs and IMPATTs State of the Art and Applications", Microwave Journal, pp. 22-28, February 1977

J. M. Osepchuk and R. W. Bierig, "Future Trends in Microwave Power Sources", IEEE Electro 76 Professional Program, pp. 1-4 May 1976

S. K. Barton, "Methods of adaptive Cancellation for Dual Polarization Satellite Systems", The Marconi Review, Volume XXXIX, No. 200, pp. 1-24, First Quarter 1976

"LSI Data Communication Device Technology Available for Commercial and Industrial Applications", Computer Design, pp. 98-101, December 1976

C. L. Cuccia, "New PCM Techniques Stress $ and Spectrum Conservation", Microwave Systems News, p. 57, January 1977

J. Collins, "SAW Devices Emerging into Commercial and Military Hardware", Microwave Systems News, pp. 15-21, January 1977

R. L. Camisa, B. F. Hitch, et al., "Microwave Integrated-Circuit MIS Varactor Phase Shifter", RCA Review, Vol. 36, p. 296, June 1975

S. Derman, "Out in Space or on the Road, Discrete Devices are on the Job", Electronic Design, No. 1, pp. 38-40, January 4, 1977

"Thin-Film SAW Units Are Very Temperature-Stable", Electronic Design, No. 1, p. 27, January 4, 1977

International Telemetry Conference Proceedings, Washington, D.C., Volume XI, 1975

D. H. Lee, K. P. Weller, and W. F. Thrower, "Passivated Millimeter-Wave Silicon IMPATT Diodes Fabricated by Ion Implementation", Proceedings of the IEEE, February 1977

R. B. Herring, "Silicon Wafer Technology - State of the Art 1976", Solid State Technology, May 1976

J. Lepoff, "How the Schottkys Detect Without DC Bias", Microwaves, pp. 44-48, February 1977

C. L. Cuccia, "Above 10 GHz Satcom Bands Spur New Earth Terminal Development", Microwave System News, pp. 37-56, March 1977

H. L. Crispin, "Earth Stations Dot the Globe", Microwave System News, pp. 30-33, March 1977

"Mini FET Amps Bandaid Systems from 4 - 18 GHz", Microwave System News, p. 73, February 1977

G. R. Davis, "Earth Stations Slim Down to Meet Grass-Root Demands", Microwaves, pp. 36-46, January 1977

"High-Performance Schottky Diodes Endure High Temperatures", Technology Utilization Officer, MSFC, Code AT01, Marshall Space Flight Center, Alabama

"European Update", Microwaves, pp. 36-46, November 1976

S. V. Bearse, "Power! GaAs FETs Star in a New Role", Microwaves, pp. 36-43, February 1977

M. E. Hines, R. S. Posner, and A. A. Sweet, "Power Amplification of Microwave FM Communication Signals Using a Phase-Locked Voltage-Tuned Oscillator", IEEE Transactions on Microwave Theory and Techniques, pp. 393-403, July 1976

D. J. Sommers, C. I. Parerd, and J. G. DiTullio, "Beam Waveguide Feed with Frequency-Reuse Diplexer for Satellite Communication Earth Station", Microwave Journal, Vol. 18, No. 11, pp. 51-58, November 1975

B. S. Skingley, "Spectrum Efficiency Key to North Sea Tropo Network", Microwaves, pp. 48-52, November 1976

"Optical Modulator Developed for High-Power IR Lasers", Electronic Design 25, December 6, 1976

R. A. Pucel, D. J. Masse, and C. F. Krumm, "Noise Performance of Gallium Arsenide Field-Effect Transistors", IEEE Journal of Solid-State Circuits, Vol. SC-11, No. 2, pp. 243-254, April 1976

B. Berson, "GaAs FETs: A Snapshot in Time", Microwave Systems News, p. 95, February 1977

C. L. Cuccia, "Communication Systems - LNA Scorecard", Micro-Wave Systems News, February 1977

E. Tahan, "Supercomponents Current and Future Trends", Microwave Journal, Vol. 19, No. 3, pp. 16-28, March 1976

"Redesigned FET Looks Promising for Audio Use", Electronic Design 5, p. 80, March 1, 1977

O. Pitzalis, Jr., "Status of Power Transistors Bipolar and Field Effect", Microwave Journal, Vol. 20, No. 2, pp. 30-61, February 1977

"4-GHz GaAs FET Amp Guarantees a 1.55 dB Noise Figure", Microwaves, March 1974

"Broad-Foil Lead Increases p-i-n Diode Switch Power", Electronic Design 5, March 1, 1975

S. V. Bearse, "Tube Manufacturers Stress Power and Efficiency Gains", Microwaves, pp. 9-10, January 1977

S. Amoroso, "Hand Held Communicator", Countermeasures, pp. 14-53, December 1976

E. D. Maloney, "Advanced Electron Tubes for Next Generation Satellite Telecommunications", Microwave Journal, Vol. 19, No. 5, pp. 49-51, May 1976

J. H. McElroy, et al., "CO_2 Laser Communication Systems for Near-Earth Space Applications", Proceedings of the IEEE, Vol. 65, No. 2, pp. 221-251, February 1977

B. Robinson, "Supercomponents - Subsystem Performance from Component Technology", Microwave Journal, pp. 29-34, May 1977

J. E. Mazo and J. Salz, "On Optical Communication via Direct Detection of Light Pulses", The Bell System Technical Journal, Vol. 55, No. 3, pp. 347-369, March 1976

"Global and Regional Monitoring from Airborne and Satellite Platforms", Optical and Quantum Electronics, Vol. 8, No. 2, pp. 185-187, March 1976

"'Recombination' Laser Readied for Space Use", Electronic Design 26, p. 20, December 20, 1976

I. Jacobs, "Lightwave Communications Passes its First Test", Bell Laboratories Record, pp. 291-297, December 1976

I. O. Shkarofsky and H. J. Moody, "Performance Characteristics of Antennas for Direct Broadcasting Satellite Systems Including Effects of Rain Depolarization", RCA Review, Vol. 37, pp. 279-288, September 1976

C. B. Rubinstein and J. O. Limb, "New Dimensions in Color Picture Coding", Bell Laboratories Record, pp. 299-303, December 1976

D. N. Kaye, "Military Turning to CMOS/SOS for Radiation-Hardened Circuitry", Electronic Design, Vol. 23, No. 7, pp. 26-28, April 1975

A. E. Atia and H. E. Williams, "Waveguide Filters for Space - Smaller, Lighter ... and Fewer Cavities", Microwave Systems News, Vol. 6, No. 4, pp. 75-77, August/September 1976

S. V. Bearse, "Semiconductor and Circuit Designers Team up to Score New Performance Highs", Microwaves, pp. 9-10, February 1977

"AM/AW-4240 Series Ultra-Low NF Amplifier 3.7 to 4.2 GHz", Data Sheet, Avantek, Inc., Santa Clara, California, September 1976

"Satellite Communications and RF Telemetry Products", Microdyne Corporation, Rockville, Maryland, 1976

"Receiving Equipment", Watkins-Johnson, Gaithersburg, Maryland

"X-Band Radar Receiver Has Everything but Bulk", Electronic Design 3, p. 13, February 1, 1977

"Schottky Diode Mixer Enables Imaging System to Operate at 90 GHz", Electronic Design 2, pp. 30-31, January 18, 1977

"Implanting Sulphur Ions Produces Super GaAs FET", Electronic Design 9, p. 111, April 26, 1977

F. R. Arams and H. C. Okean, "State-of-Art Low Noise Receivers For Microwave and Millimeter Communications", Proceedings of The 1976 National Telecommunications Conference, pp. 48.5-1-4

F. G. Lynn, "Impatts Surge to Higher Powers and Frequencies", Microwaves, pp. 9-16, April 1977

"FET Generates 1W at 8 GHz", Electronics, p. 151, May 26, 1977

A. Lekholm, F. Sellberg, P. Weissglas, and G. Anderson, "A Reliable Fifteen-Percent-Efficiency Silicon Double-Drift-Region IMPATT Diode.", Proceedings of the IEEE, pp. 1613-1615, November 1975

"Airborne Electronically Steerable Phased Array, Final Report", Texas Instruments Incorporated, April 1976

"Performance Specification for Services Via The Tracking and Data Relay Satellite System", S-805-1, Goddard Space Flight Center, Maryland, November 1976

V. Sokolov and H. Q. Tserng, "Characteristics of Solid State Microwave Power Amplifiers for Telecommunication Applications", Proceedings of The 1976 National Telecommunications Conference, pp. 48.2-1-5

R. Hartman, "The Solid State of Microwaves", Military Electronics/ Countermeasures, pp. 32-33, August 1977

J. Steinkamp, "Trends in Semiconductor Application for Microwave Communication Systems", Microwave Journal, p. 57, March 1977

R. S. Napier and C. E. Foster, "18 to 40 GHz Broadband Frequency Synthesizer Technique", Microwave Journal, pp. 40-47, April 1977

"TWT for Military Satellite Systems", Military Electronics/ Countermeasures, p. 44, August 1977

"TRW Builds 5-GHz GaAs Modem Chips", Electronics, p. 40, May 12, 1977

G. Marzocchi, "Choice of Semiconductors for FM High-Capacity Radio Link", Microwave Journal, pp. 58-60, March 1977

H. J. Sigg, G. D. Vendelin, T. P. Cauge, and J. Kocsis, "D-MOS Transistor for Microwave Applications", IEEE Transactions on Electron Devices, pp. 45-53, January 1972

F. Ivanek, "Extending Direct MW Generation and Amplification Toward Higher Frequencies", Microwave Journal, pp. 60-61, March 1977

J. A. Giordmaine, "Solid-State Electronics: Scientific Basis for Future Advances", Science, Vol. 195, pp. 1235-1240, March 1977

R. Fekete, "11 GHz 2W Solid-State Power Amplifier for LOS Microwave Radio Equipment", Proceedings of The 1976 National Telecommunications Conference, pp. 48.3-1-4

R. Van Tuyl and C. Liechti, "Gallium Arsenide Spawns Speed", IEEE Spectrum, pp. 41-47, March 1977

"High Power/High Efficiency Read Profile Impatt Diodes 9-11 GHz", Raytheon Corporation, Waltham, Massachusetts, February 1977

"Try Mospower FET's in your Next Broadband Driver", Siliconix Incorporated, August 1976

T. J. Rodgers and J. D. Meindl, "VMOS: High Speed TTL Compatible MOS Logic", IEEE Journal of Solid State Circuits, pp. 239-249, October 1974

"High Power/High Efficiency Read Profile Impatt Diodes 12-14 GHz", Raytheon Corporation, Waltham, Massachusetts, January 1977

L. Altman and L. Mattera, "Will V-MOS Make it in the Power Market", Electronics, p. 84, September 15, 1977

"Motorola, Fairchild to Bring out V-MOS Power Parts", Electronics, p. 34, September 15, 1977

"New TWT Cathodes Look Promising", Electronics, pp. 34-36, September 30, 1976

"Microwave Signal Sources Catalog", Varian Associates, Palo Alto, California

"Extended Interaction Oscillators", Data Sheet, Varian Associates of Canada, LTD., Ontario, Canada, March 1974

A. W. Love, "Some Highlights in Reflector Antenna Development, Radio Science, pp. 671-684, August-September 1976

"Antenna Update", Microwave Systems News, pp. 84-85, March 1977

R. A. Semplak, "100-GHz Measurements on a Multiple-Beam Offset Antenna", The Bell System Technical Journal, pp. 385-398, March 1977

D. Hackmeister, "Radio Receiver Reaches 670 GHz with 'Quasioptics' and New Mixer Mount", Electronic Design 23, pp. 26-27, November 8, 1977

H. W. Ehrenspeck and J. A. Strom, "Short-Backfire Antenna - A Highly Efficient Array Element", Microwave Journal, pp. 47-49, May 1977

R. W. Kreutel, Jr., et al., "Antenna Technology for Frequency Reuse Satellite Communications", Proceedings of the IEEE, Vol. 65, No. 3, pp. 370-377, March 1977

F. Dintelmann and F. Rucker, "Analysis of an XPD Event at 30 GHz Measured with ATS-6", Proceedings of the IEEE, Vol. 65, No. 3, pp. 477-479, March 1977

K. J. Russell and R. S. Harp, "A Multistage High-Power Solid-State X-Band Amplifier", 1978 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pp. 166-167, February 1978

J. M. Schellenberg and Marvin Cohn, "A Wideband Radial Power Combiner for FET Amplifiers", 1978 IEEE International Solid-State Circuits Conference Digest of Technical Papers, pp. 164-165, February 1978

6. PREPROCESSING ELEMENTS

Preprocessing elements are defined herein to include the elements identified in Figure 6-1. Thus the preprocessing elements consist of the data storage and data manipulation (hardware and software) elements that take either the pre-detection or post-detection serial bit stream and detect (as necessary), buffer or store, demultiplex, and reformat into a parallel, digital output with calibration. Figure 6-2 is a block diagram of the data transfer, preprocessing, and processing elements as discussed in this report. Synchronization, decoding, decommutation/demultiplexing, and wideband recording technologies are presented in this section. The preprocessing functions use computers for both control and data reduction. The discussion of the processing elements is presented in Section 7. Similarly, software is discussed in Section II.

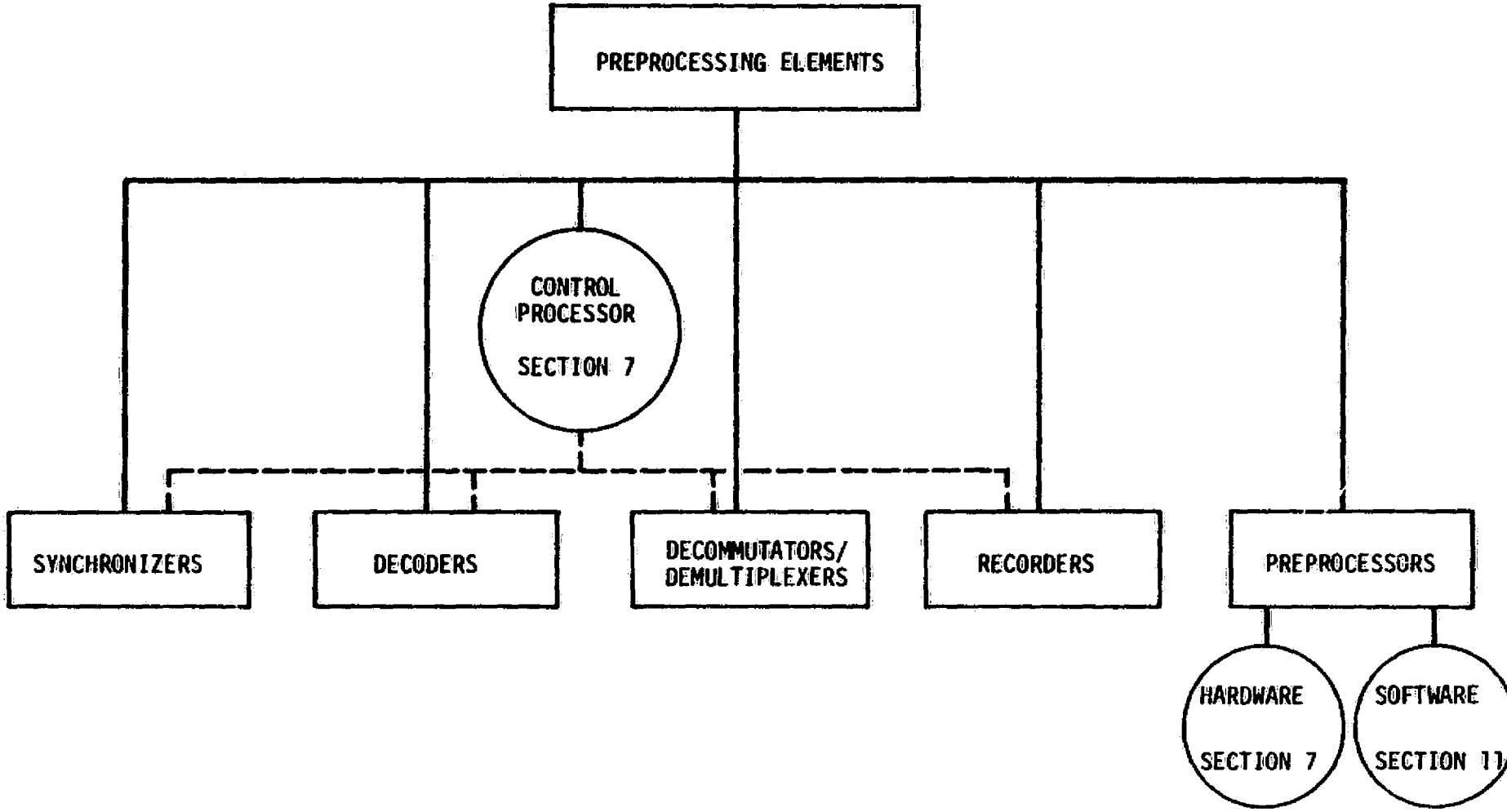


FIGURE 6-1. PREPROCESSING ELEMENTS

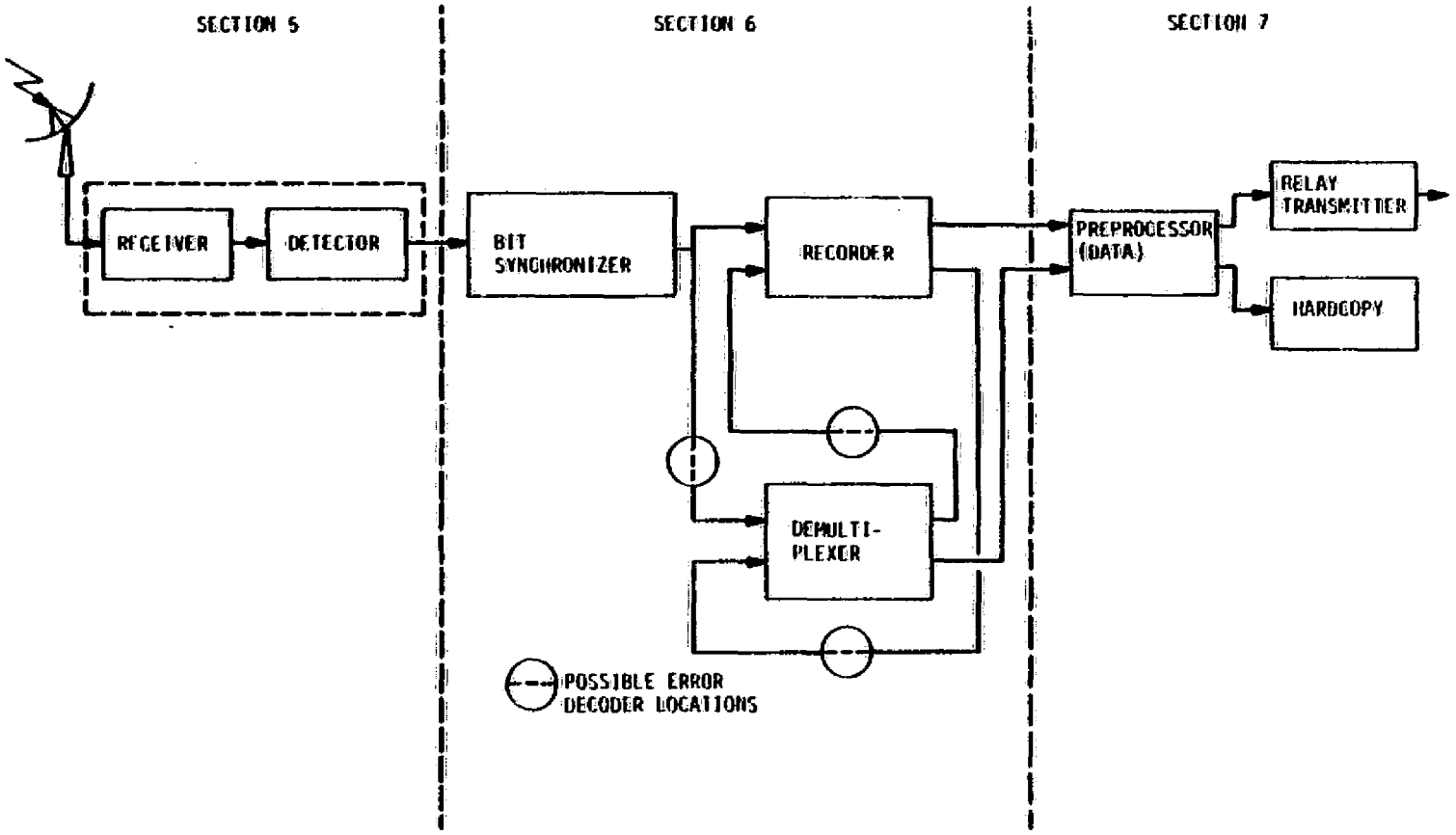


FIGURE 6-2. DATA TRANSFER, PREPROCESSING, AND PROCESSING ELEMENTS

6.1 BIT SYNCHRONIZER ELEMENTS

This section addresses the technology for bit synchronizers as used for ground station preprocessing. The bit synchronizer is normally interfaced between the output of the detector and the input of the demultiplexer or recorder, as illustrated in Figure 6-2. The bit synchronizer input is normally a noisy PCM signal that is typically in either bi-phase or NRZ format. The bit synchronizer uses some type of phase-lock technique to lock an internal oscillator onto the incoming signal to derive an output clock that is in phase with the serial binary output. Additionally, the bit synchronizer generally employs some type of filter such as a digital matched filter (or integrate and dump) to improve the signal quality.

Bit synchronizer hardware normally falls into two divisions: those units operating at data rates below 20 Mbits/sec and those operating above 50 Mbits/sec. The units operating below 20 Mbits/sec typically have variable-input data rates that are front panel or keyboard programmable. The units operating above 50 Mbits/sec are usually custom designed for specific data rates that can be varied only within a limited range.

6.1.1 State of the Art in Bit Synchronizers

Currently available bit synchronizers perform several signal conditioning functions in the ground receiving station. The data coming from the detector are subject to a timing ambiguity due to phase jitter and clock frequency instability. The bit synchronizer employs a phase-locked loop to track the input data stream and in turn sends out the data at a rate that is synchronized either with an internal clock or more likely with the station clock. Additionally, the amplitude level of the data is stabilized so that wide variations in level do not affect the other preprocessing elements.

The characteristics for several state-of-the-art data bit synchronizers are listed in Table 6.1.1-1. The EMR model and the IED model are both typical of the lower-rate programmable synchronizers. They cover variable data rates over widely programmable ranges and accept a number of coding formats.

The Harris model is of more interest due to the high data rates involved. This synchronizer will operate at a rate of 120 or 240 Mbits/sec ($\pm$5%) for BPSK or at 240 or 480 Mbits/sec ($\pm$5%) for QPSK. Additionally, the hardware can be modified to cover a $\pm$10% range with some performance degradation. Performance at these high data rates is exceptional, with bit error degradation with 1 dB of the theoretical value of BPSK. The performance characteristics of the synchronizer minimizes degradation due to input level variations, bit-rate static changes, phase jitter, baseline variations, bandlimiting, shifts in transition density, and noise. Bit slippage is less than 1×10^{-7} for a signal-to-noise ratio of 2 to 12 dB. The output of the synchronizer is compatible with emitter-coupled logic.

TABLE 6.1.1-1. CHARACTERISTICS FOR SEVERAL STATE-OF-THE-ART BIT SYNCHRONIZERS

MODEL	DATA RATE (bits/sec)	CODES	LOOP BANDWIDTH (%)	TRACKING RANGE (%)	BIT ERROR PERFORMANCE
EMR 720-02	1 to 5M	NRZ-L NRZ-M NRZ-S Biϕ-L Biϕ-M Biϕ-S RZ DM-M DM-S	0.1 0.3 1.0	1.0 3.0 10.0	Within 1 dB of Theoretical up to 2.5 MHz
IED d/pad one	100 to 1M	NRZ Biϕ	0.05 to 10	--	--
HARRIS (for NASA)	120M/240M	NRZ-L	0.1 to 2	±5	Within 1 dB of Theoretical

6.1.2 Trends and Projected Developments in Bit Synchronizers

The general trend for satellite data communications is toward higher data rates, as discussed elsewhere in this report. The lower-bit-rate synchronizers will change very little in capability during the 1980-1985 timeframe. The major changes in these units will be a shift to microprocessor control and a resulting increase in programming ability and status monitoring. The low-data-rate models now offering microprocessor control perform such functions as self diagnosis and testing, bit-error-rate calculations, and CRT display of system parameters, in addition to the bit, frame, and subframe synchronization. These functions and many more will be available as more manufacturers introduce a line of microprocessor-controlled synchronizers.

The high-data-rate synchronizers will be of more importance for ground station use in the 1980-1985 timeframe. The implementation of the Tracking and Data Relay Satellite System (TDRSS) with its high-data-rate capabilities (up to 300 Mbits/sec) and the evolving of high-data-rate orbital systems such as Space Shuttle and Landsat-D will cause the need for high-rate-data synchronizers. Custom-built models operating at 240 Mbits/sec for BPSK are available currently, and laboratory models approaching 1 Gbit/sec have been implemented by several researchers. Projections for data rates in the 1985 timeframe extend to 1 Gbit/sec and above. Custom-built bit synchronizers will be capable of handling these 1-Gbit/sec rates, but standard off-the-shelf commercial units will not provide rates in excess of 50 Mbits/sec in this period. The custom-built, high-rate synchronizers will rely on very-high-speed logic such as Static Induction Transistor logic, Shottky GaAs MESFET logic, or TED GaAs logic families. The general characteristics of the 1985 bit synchronizer will not differ greatly from the 240-Mbit/sec models now available. The major difference will be in data-rate capacity.

6.2 SIGNAL DECODER ELEMENTS

This subsection addresses the technology for error-correcting decoders as are used on current and planned NASA missions. The decoder can be placed in the ground station data link at several locations, as shown in the block diagram of Figure 6-2. Error-correcting codes are discussed in Subsection 5.6.1.3, the decoding hardware is discussed in this section.

Current coding techniques generally use either block codes (i.e., BCH, Reed-Soloman) or convolutional codes. In some applications, concatenated (multiple) codes are used, with a combination of block and convolutional codes for high data reliability or pseudo-random noise (PN) and convolutional codes for unique characteristics (as for TDRSS ranging and tracking). The shift in emphasis is toward convolutional codes, and they will continue to gain importance with the implementation of the Space Shuttle and the TDRSS in the early 1980's.

6.2.1 State of the Art in Decoders

Today's error-correction decoders are available in a wide range of types covering assorted codes and bit rates. These decoders are available for each of the many block codes and for the several types of convolutional codes. The more popular types of convolutional decoders are the sequential decoders, maximum-likelihood (Viterbi) decoders, and convolutional binary feedback decoders. Figure 6.2.1-1 presents the performance characteristics for several types of decoders, both block and convolutional. The sequential decoders and the Viterbi (coding rate 1/3) decoders both perform better than the other types of Convolutional decoders. The Viterbi (rate 1/2) decoder performs better than the block decoders or the convolutional feedback decoder, is simpler than the sequential decoder, and requires less channel bandwidth than the Viterbi rate 1/3 decoder.

The implementation of the TDRSS to relay spacecraft data through a geosynchronous satellite to Earth presents several unique problems, one of which is a restriction on available signal power, especially for wideband links. To maintain a reasonable bit error rate (BER) while working in a power-limited environment, convolutional coding was chosen because of its significant coding gain over block codes. Additionally, the convolutional coding system has a simple spaceborne encoder design, and it maintains a relative simplicity and wide availability for ground station decoders. The choice of convolutional coding for the TDRSS and Space Shuttle effectively sets the coding specifications for spacecraft during the 1980's. For this reason, convolutional decoders will be the main area of discussion for this subsection.

The two major types of convolutional decoders are the Viterbi decoders and the sequential decoders. Both of these have certain unique characteristics, and the selection of the decoder type must be based on the application and the performance requirement. Table 6.2.1-1 presents a list of relevant characteristics that should be considered in selecting a decoder. Each decoder has a certain amount of delay associated with it, and a buffer is supplied to hold the data prior to processing. In

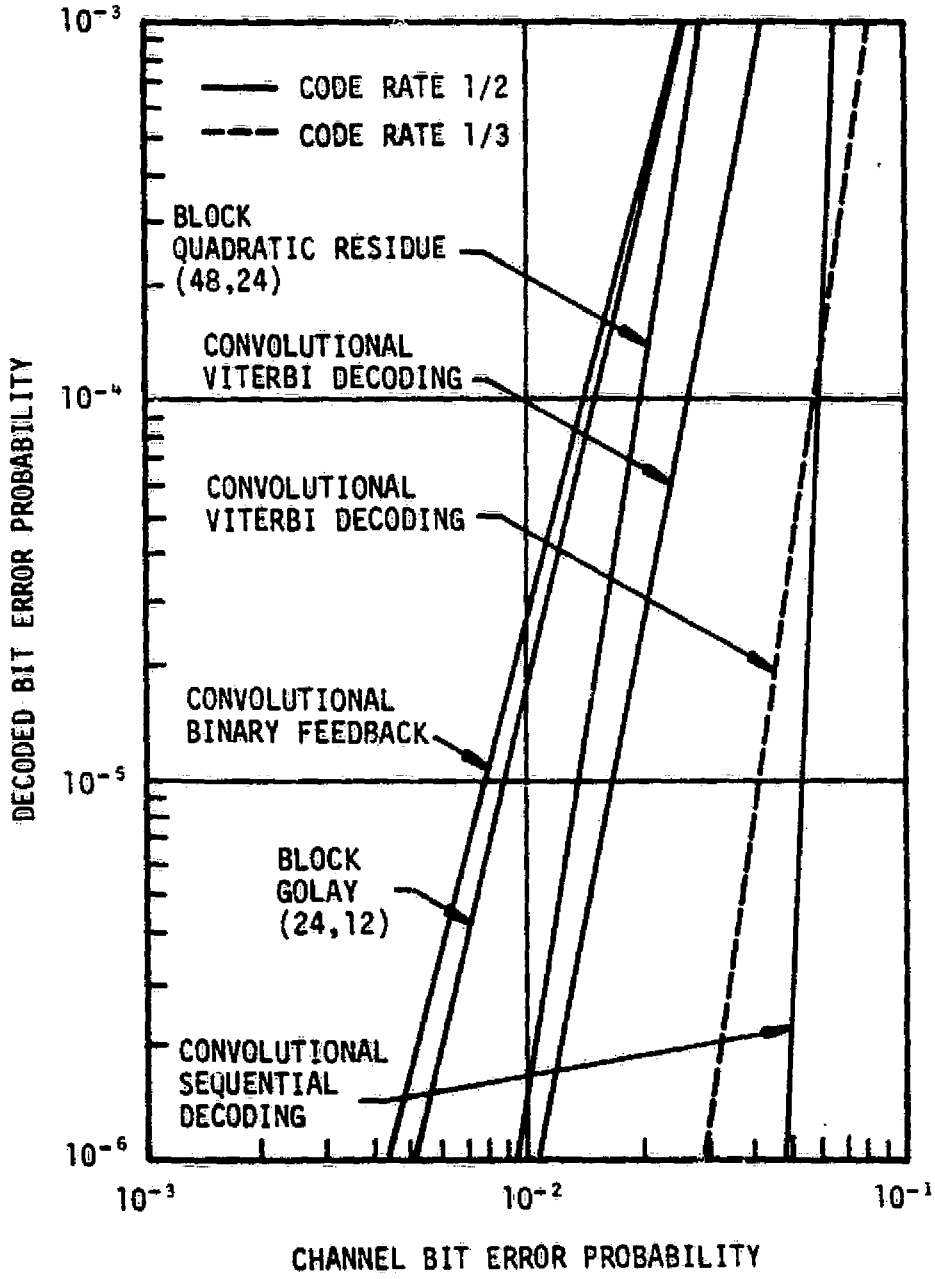


FIGURE 6.2.1-1. DECODING PERFORMANCE FOR SEVERAL STATE OF THE ART DECODER TECHNIQUES. (Ref. 6-1)

TABLE 6.2.1-1. CHARACTERISTICS FOR SEVERAL TYPES OF CONVOLUTIONAL DECODERS

	DECODER TYPE	
CHARACTERISTIC	MAXIMUM LIKELIHOOD	SEQUENTIAL
Overflow Errors	Short bursts of 10 to 20 bits	Long bursts of several thousand bits with error rate of the raw channel
Achievable Coding Gain for BPSK	7.7 dB at 10^{-6} BER with R = 1/3 6.5 dB at 10^{-6} BER with R = 1/2	8.8 dB at 10^{-6} BER with R = 1/2
Decoder Delay	64 bits	200 to 65,000 bits
Tailoff Degradation	Minimum amount	Severe degradation due to long bit delay sequence
Coding Rates	1/2, 1/3, others	Limited to 1/2; other rates require complex hardware
Quantization	Hard decision or soft quantization	Limited to hard decision
Type of Applications for Optimum Performance	Channels with frequent errors and short bursts	Channels with occasional errors and long bursts

the event of input data with an abnormally high error rate, the decoder is unable to process the data at a sufficient speed to avoid a buffer overflow. If a buffer overflow occurs, the maximum-likelihood decoder will produce data with short-burst errors of 10 to 20 bits. The sequential decoder produces long bursts of several thousand bits with the error rate equivalent to that of the input channel.

The amount of decoder delay is important, especially for high-speed data. The Viterbi decoder has a relatively short delay, typically less than 64 bits. The sequential decoder has a delay that varies from approximately 200 bits to as much as 65,000 bits, with the longer delays providing the better error correction performance. Systems operating with switched inputs, such as time division multiple access, lose a continuity of the error code when switched and experience tailoff degradation. The severity of this degradation is proportional to the decoder delay, and the longer delay lengths of sequential decoders make them especially susceptible to this problem.

The amount of coding gain that can be achieved with a particular system is likewise an important characteristic. The achievable coding gain is a function of the output error rate, and for purposes of comparison, a 10^{-6} BER is chosen. The sequential decoder provides the highest coding gain, typically 8.8 dB with a rate 1/2 code and BPSK modulation. The maximum-likihood decoder provides a 7.7-dB coding gain at a coding rate of 1/3 and a 6.5-dB coding gain at a rate of 1/2. The Viterbi decoders are easily adaptable to coding rates of 1/2, 1/3, and other rates, but the sequential decoders are limited to rate 1/2 because of greatly increased hardware complexity at the other rates. Likewise, the sequential decoder is limited to hard-decision decoding, while the maximum-likelihood decoders can be adapted to either hard-decision or soft quantization. The maximum-likelihood decoders are better suited for channels having frequent errors in the form of short bursts. The sequential decoders are better suited for channels with occasional errors that consist of long bursts.

Convolutional decoders operating at data rates of up to 100 kbits/sec are quite common, and rates well over a megabit are widely available. Viterbi decoders with data rates of 10 Mbits/sec are generally considered to be the maximum value for a standard "off-the-shelf" configuration. Higher data rates of up to 35 Mbits/sec can be achieved using ECL, but these devices are not available as standard equipment at present. A 50-Mbit/sec maximum-likelihood decoder has been built in the laboratory, but constraint length, complexity, and long delay times limit its applications. Sequential decoders likewise are common at data rates up to approximately 10 Mbits/sec. Several commercially available sequential decoders handle data rates of up to 40 Mbits/sec, but operation becomes marginal at these higher rates.

The decoder design chosen to handle the 50-Mbit/sec downlink from the Space Shuttle is a system using five 10-Mbit/sec decoders in parallel. Figure 6.2.1-2 is a functional block diagram of the 50-Mbit/sec decoder system used for Shuttle data at the TDRSS ground station. The data enter the bit synchronizer and are routed through the demultiplexer with each consecutive bit going to a consecutive decoder. This method allows a more conventional technology to be applied to the high data rates. The parallel system does exhibit a lower reliability than would be obtained with a single unit, but this is corrected by placing spares into the stack and incorporating an automatic fault detection scheme.

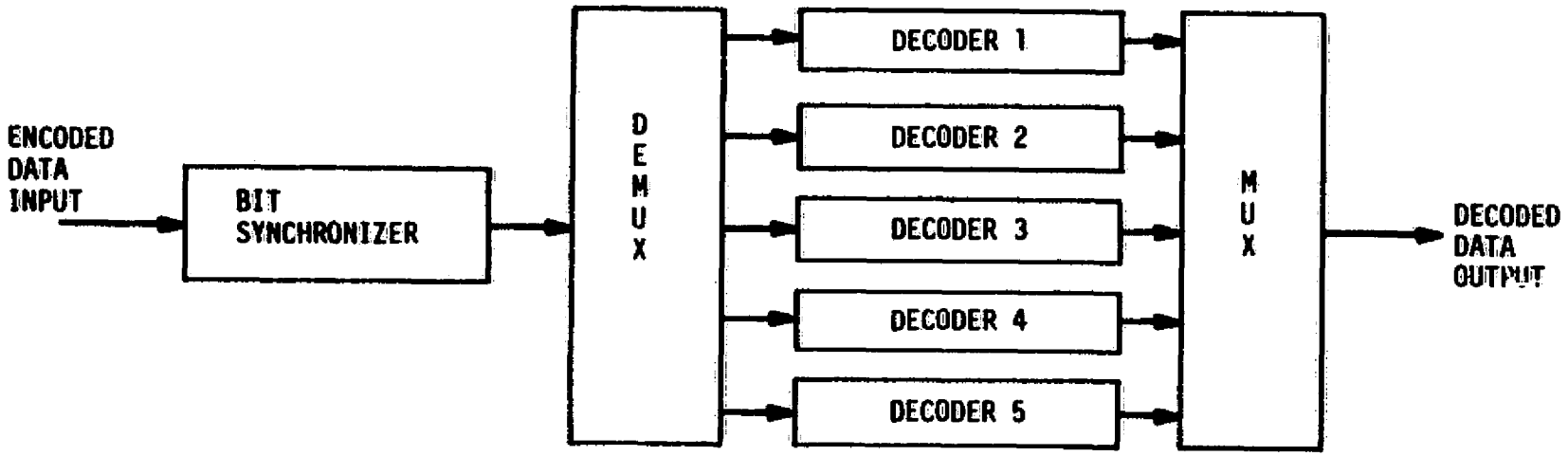


FIGURE 6.2.1-2. FUNCTIONAL CONFIGURATION FOR THE SPACE SHUTTLE 50 Mbit/sec DECODER

6.2.2 Trends and Projected Developments in Decoders

The trends in error-correcting decoders will generally be toward an increase in data rates and quality of data and a decrease in power, size, and weight requirements. Table 6.2.2-1 presents the state of the art and projected characteristics for a typical 10-Mbit/sec maximum-likelihood decoder. The state-of-the-art unit is based on a commercially available design with mostly TTL integrated circuits. The decoder designs of 1980 will use some MOS technology and will exhibit higher speeds with decreased power and size. By 1985, decoders using GaAs and other GHz logic technologies (see Section 1) will be commercially available, providing data rates of up to 100 Mbits/sec for a single decoder unit while reducing power and size requirements. In each case, parallel processing can be applied, allowing the decoding of higher data rates. Using parallel processing, 150 Mbits/sec can be achieved in 1980 and a 1-Gbit/sec data rate could be processed by 1985.

Figure 6.2.2-1 reflects the trends in decoder power requirements for the 1977-1985 timeframe. Power will decrease from the 225-W requirement of the state-of-the-art 10-Mbit/sec decoder to slightly over 100 W for a 100-Mbit/sec decoder by 1985. This reduction is mainly credited to lower power logic families. Figure 6.2.2-2 presents the trends in decoder size. Large-scale integration and lower power dissipation will allow the 1985 size to decrease to approximately one-third of the 1977 size. Decoder data rates will increase as shown in Figure 6.2.2-3. Higher-speed logic will mainly be responsible for the increase in the data rate of a single decoder. The trend for parallel decoder banks is also shown on the chart.

TABLE 6.2.2-1. STATE OF THE ART AND PROJECTED CHARACTERISTICS FOR DECODERS

CHARACTERISTIC*	STATE-OF-THE-ART	PROJECTED 1980	PROJECTED 1985
Data Bit Rate (Mbits/sec)	10	25	100
Data Bit Rate (Mbits/sec) (Parallel Processing)	50	150	1,000
Power (W)	225	170	110
Size (in^3)	900	630	315
Weight (lb)	25	17	9
Cost ($)	12.7k	TBD	TBD

*Characteristics for single decoder except when noted

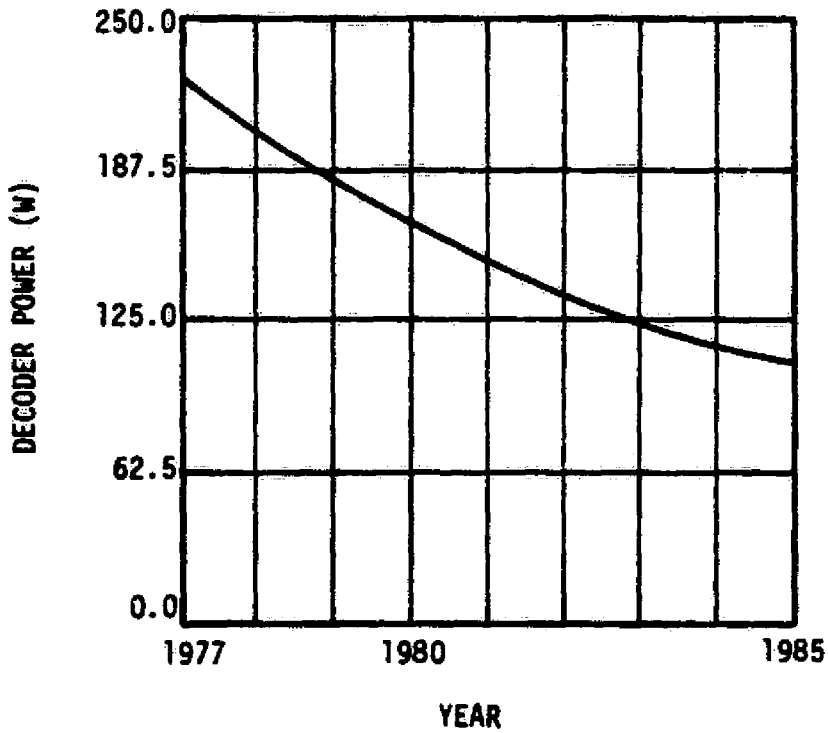


FIGURE 6.2.2-1. TRENDS IN DECODER POWER REQUIREMENTS

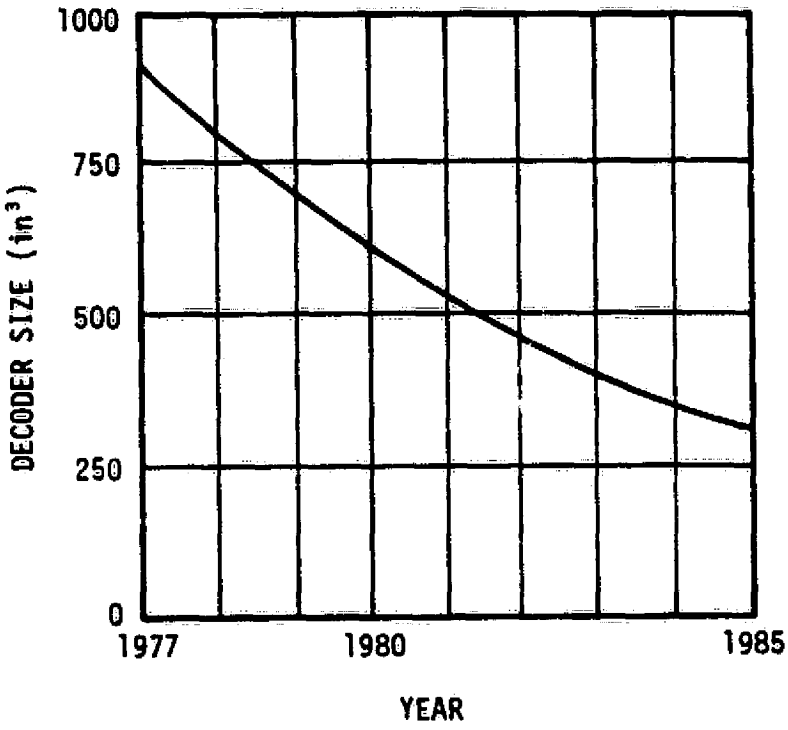


FIGURE 6.2.2-2. TRENDS IN DECODER SIZE

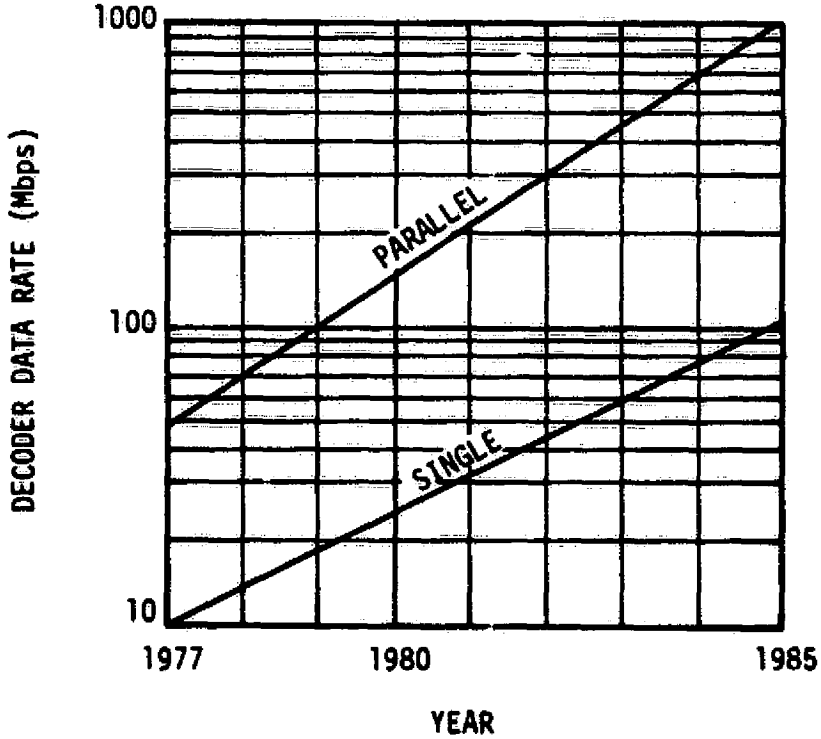


FIGURE 6.2.2-3. TRENDS IN DECODER DATA RATES

6.3 DEMULTIPLEXER/DECOMMUTATOR ELEMENTS

Demultiplexer/decommutator elements fit into the preprocessing network as shown in Figure 6-2. These elements are divided into either analog or digital classifications, as shown in Figure 6.3-1. Analog demultiplexers are further subdivided into frequency-division multiplexing (FDM) and time-division multiplexing (TDM) systems. Since analog TDM has been replaced by digital PCM for space data handling applications, advances in analog TDM decommutator technology are not reported.

In general, the technology for decommutators and demultiplexers is comparable, with the technology for multiplexers as reported in Subsections 4.1.1 and 9.1.1.2, since the same system technology (e.g., data rates, bandwidths, and codes) applies in both areas. In fact, decommutator technology should be ahead of spacecraft multiplexer technology because of less restrictions on size, power consumption, weight, etc.

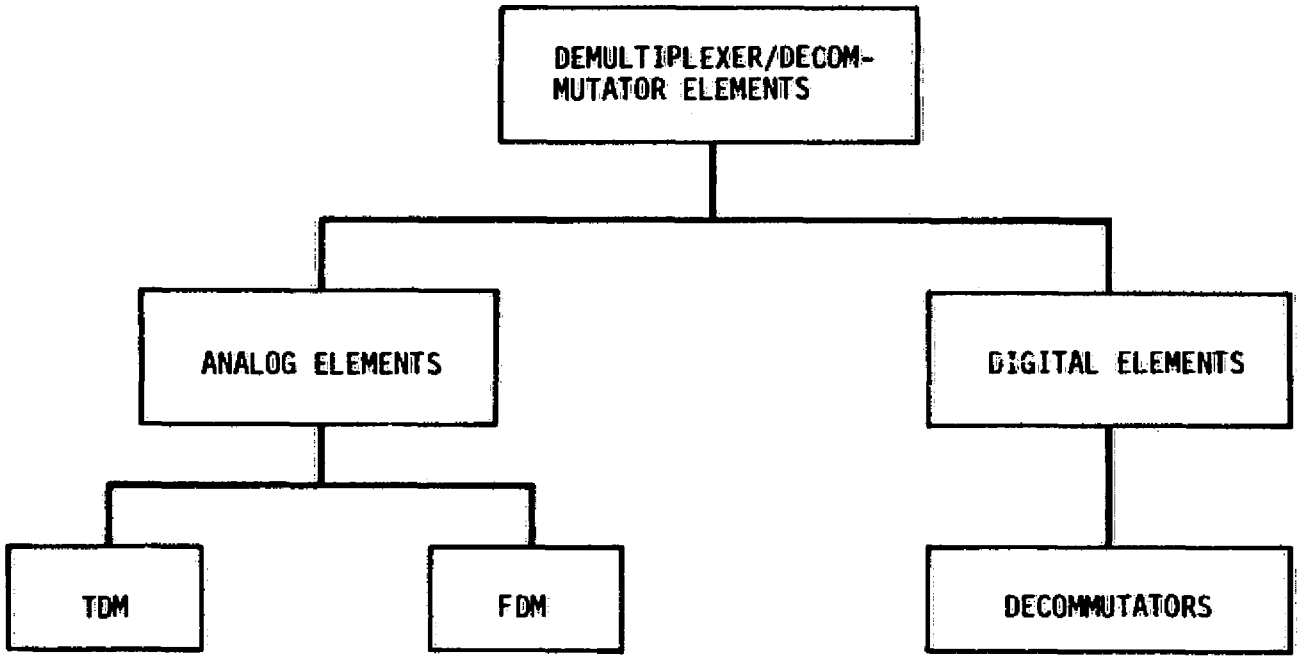


FIGURE 6.3-1. DEMULTIPLEXER/DECOMMUTATOR ELEMENTS

6.3.1 State of the Art in Demultiplexer/Decommutator Elements

FDM signals are defined by IRIG 106-75, "Telemetry Standards". The development of FDM decommutators in recent years has been toward smaller, more universal, plug-in type discriminators with low power consumption. However, until recently very little has been done toward increasing the number of data channels or data capacity. As with other areas of communication electronics, the development of MSI and LSI circuits has done much to reduce the size and power consumption of these units. Because of the trend toward higher data rates and all-digital systems, there has been very little new development in FDM demultiplexers.

The number of FDM wideband channels was increased by IRIG 106-77 to provide additional channels above the former constant-bandwidth and proportional-bandwidth channels. The number of narrowband constant-bandwidth channels did not increase, however. Systems requiring a larger number of narrowband channels are required to use nonstandard channels to accommodate the data needs.

Digital demultiplexers represent the area of demultiplexing where the greatest changes resulting from technology improvements have been made in recent years. Demultiplexers available today on a commercial basis tend to be capable of handling data rates on the order of 2 to 5 Mbits/sec, with some units available at 10 Mbits/sec on a more-or-less custom basis. Current technology permits decommutation hardware capable of handling hundreds of megabits, and a few systems are in existence that will handle these rates. As discussed in Subsection 9.1.1.2, the terrestrial telephone network is using a digital coaxial cable system operating at 274-Mbit/sec rates. Additionally, the TDRSS ground station, which is being built by Harris under a subcontract from TRW, will be capable of handling data rates of up to 300 Mbits/sec.

The Air Force has an operational laboratory model of a 1-Gbit/sec digital data system that includes a six-channel demultiplexer. The system uses custom-designed integrated circuits.

6.3.2 Trends and Projected Developments in Demultiplexer/ Decommutator Elements

Improvements in LSI and standardization of high-data-rate systems will allow a reduction in size and power consumption and an increase in capabilities. INTELSAT has developed a set of LSI circuits that are available to users to provide PCM encoding/decoding, convolutional coding/decoding, transmitter synchronization, and bit timing recovery for their system. High-speed logic circuits, such as the Plessey 1.2-GHz programmable divider, have become common.

Increased flexibilities in decommutators can be expected through 1985 as a result of the availability of low-cost memory and microprocessors. The "decom" boxes will contain more internal intelligence and will become multiported to accommodate distributed processing of high-data-rate signals. More digital signal processing will be employed in the signal detection and synchronization areas. Improvements will be made in signal processing performance, and increased throughput (data rate), through the use of distributed processing, and improved human factors for units can be expected. Increased development of LSI/VLSI high-speed/low-power silicon and GaAs logic technologies with their reduced propagation delays will significantly increase demultiplexer data rates by 1985.

The multiplexers used in future space-to-ground data links are expected to be digital. Two types of digital demultiplexers are foreseen: general-purpose and unique/dedicated, with the general-purpose types being restricted to the lower data rates and the unique/dedicated types being built for special-purpose systems or experiments such as the Thematic Mapper.

The technology will support hundreds of megabits; however, standard commercial products for 1980 will be of the general-purpose type and capable of handling 10 to 12 Mbits/sec, with increases to 35 to 40 Mbits/sec for 1985. At the higher data rates, custom demultiplexers will be built using either ECL, GaAs MESFETS, or other high-speed logic technologies that become available.

Decommutators will have more output ports to accommodate distributed processing of high-data-rate signals. They will also be more adaptive in an attempt to overcome nonstationary noise and signal conditions. Concatenated coding will be implemented to correct errors from random noise, burst errors, and signal dropout. The error probabilities will be within 1 dB of theoretical for data rates up to 5 Mbits/sec, within 2 dB for 5 to 20 Mbits/sec, and within 3 dB for 20- to 50-Mbit/sec signals by 1985.

6.4 WIDEBAND RECORDING ELEMENTS

This section addresses the technology for instrumentation type tape recorders such as those found in telemetry ground receiving stations for recording wideband analog and digital (PCM) data. Section 3.1 addresses the technology for using magnetic tape recorders for spacecraft data storage, and Subsection 7.2.5 examines the technology for data processing magnetic tape recorders. Most of the discussion in Section 3.1 is applicable to ground instrumentation recorders, particularly the projections of Subsection 3.1.3.

Magnetic tape recorders have been used in the instrumentation area for many years. During this time, major progress has been made in improving the performance through advances in the recording heads, the electronics, the drives, data encoding, and the media (i.e., the tape). Also, standards have evolved for tape speeds, tape widths, data bandwidths, etc.

Currently, most of the new developments in instrumentation recorders are directed toward achieving higher data rates and greater storage densities in high-rate digital recording systems.

6.4.1 State of the Art in Wideband Magnetic Tape Recorders

Today's instrumentation recorders are available in a wide variety of standard configurations as defined in Table 6.4.1-1. The recorders listed in this table are the fixed-head instrumentation type recorders such as those manufactured by Ampex and Bell & Howell. Fixed-head recorder designs have progressed from 14 tracks/in. in the 1960's to 42 tracks/in. at present. Frequency responses of 2 MHz are standard, and experimental systems are being tested with frequency responses of 4 MHz at 120 in/sec. Several companies are currently developing recorders that use 2-in.-wide tapes with up to 84 tracks; however, these recorders are not yet commercially available.

According to Bessette (Ref. 6-2), fixed-head recorders currently achieve 50% of the theoretical potential number of bits per inch (80 kbits/in) and 10% of the potential number of heads per inch (1,000/in.). This implies that fixed-head recorders achieve 5% (10% of 50%) of the theoretical data packing density (5×10^7 bits/in^2) using currently available media. These figures indicate that future improvements in recording density will be achieved by increasing track density rather than attempting to increase bit density.

Increasing the number of data tracks also increases the total input data rate for digital applications. However, for analog recorders, the bandwidth can only be increased by increasing in-track response either by increasing the data density or by moving the tape at higher speed. Factors that affect the response of a track are the tape magnetic particle size (12 μin. typical), the head gap, and the tape speed. The small gaps required for higher-frequency response require shallower pole face depths, which are difficult to achieve and have shorter lives. New methods of tape lubrication are being developed to improve the head life. Developments in magnetic materials to allow smaller particle size and thinner oxide coating are also occurring.

In addition to the "standard" instrumentation recorder configurations, there are several other types of instrumentation recorders.

TABLE 6.4.1-1. STANDARD INSTRUMENTATION RECORDER CONFIGURATION

Tape Speeds	15/16 to 240 in/sec
Tracks	7, 8, 14, 16, 28, 32, or 42
Tape Widths	1/2, 1 in.
Reel Size	8, 10 1/2, 12 1/2, 14, 15, and 16 in.
Record Methods	FM, direct, digital high density
Bandwidth/Data Rates (per track)	
FM	dc to 600 kHz (at 120 in/sec)
Direct	100 Hz to 2 MHz (at 120 in/sec)
Digital H.D.	to 3.5 Mbps (at 120 in/sec)

One is the Newell drive type recorder manufactured by Emerson. The other type is the rotary-head recorder manufactured by RCA, Ampex, and Echo Science.

The Emerson TITAN recorder features 42 tracks on a 1-in.-wide tape with tape speeds of up to 600 in/sec. The TITAN will record digital data at up to 10 Mbits/sec/track.

Rotary-head recorders offer a number of advantages, including: higher data storage density, rapid start/stop, and high rate at low tape speed. Because of these advantages, the rotary-head type recorder will find increased application for instrumentation recording in the future.

The RCA VERSABIT 200 is representative of the state of the art in rotary-head recorder designs. This system records 20 Mbits/sec on a standard 2-in.-wide television tape at a tape speed of only 9.53 in/sec. The VERSABIT 200 records at an in-track density of 10,000 bpi at 133 tpi. RCA is currently developing a double-density system with a track density of 266 tpi.

6.4.2 Trends and Projected Developments in Wideband Magnetic Tape Recorders

The trends in fixed-head instrumentation recording systems are toward wider tapes and more tracks per inch of tape. This results in higher data rates for high-density digital recording and more channels for analog recording. As a result of the trend toward digital instrumentation and communication systems, it is unlikely that analog recording will still be in use in the 1980 - 1985 timeframe for ground-based data handling systems.

Current equipment under development includes a Bell & Howell recorder being built for Northrop. The unit is to be incorporated into an Air Force system that has 84 tracks on a 2-in.-wide tape and records at an input data rate of 270 Mbits/sec. RCA has recently demonstrated a laboratory model of a 240-Mbit/sec recorder with 120 data tracks. These recorders are forerunners of the next-generation commercial fixed-head instrumentation tape systems.

Rotary-head recorders will find increased use in systems that utilize instrumentation recorders to provide temporary buffering as well as permanent data storage where frequent start/stop/replay operations are required such as on-line processing of satellite data.

Trends in rotary-head recorder designs are toward higher data rates using multiple heads, more tracks per inch of tape, and tape head designs that reduce head wear. As previously mentioned in Section 3.1, IBM's recently announced 3850 Mass Storage System employs a helical scan rotary-head system that "flies" across a 2.7-in.-wide tape with the head at 10 μin. from the tape to minimize head wear.

A number of improvements in both analog and digital recorders are within current technological capabilities. Improved tape handling will be realized through use of vacuum chambers and zero loop configurations by 1980. Improved maintenance and calibration will result from the use of microcomputer interfaces by 1985.

Analog recorders in the early 1980's will accommodate a 4-MHz bandwidth at a tape speed of 120 in/sec with up to 48 tracks/in. on either 1-in.- or 2-in.-wide tapes. The heavy trend toward digital recording places some doubt on the use of wideband analog recorders by 1985. If analog recorders are still in use by then, the bandwidth should approach 6 MHz.

The use of digital recorders will increase and very-high-density digital recorders with throughput rates of near 1 Gbit/sec at densities of 10^7 bits/in^2 will become available in the 1980-1985 timeframe. The projections presented in Figures 3.1.3-1, 3.1.3-2, and 3.1.3-3 for aerospace instrumentation recorders also apply to ground instrumentation recorders.

A number of other projections foreseen in the 1980-1985 timeframe include a move from longitudinal to rotary, helical, or transverse heads to permit the use of very narrow tracks without excessive costs; rapid start-stop capability along with continuously variable speeds; and improved reliability and head lifetimes through improved head designs, tape-lubrication, and other techniques.

REFERENCES - SECTION 6

6-1 G. K. Huth, "Optimization of Coded Spread Spectrum System Performance", IEEE Transactions on Communications, pp. 763-770, August 1977

6-2 O. E. Bessette, "Recent Advances in Magnetic Recording of Digital Data", International Telemetry Conference, Los Angeles, California, September 28-30, 1976

BIBLIOGRAPHY - SECTION 6

W. W. Wu, "New Convolutional Codes - Part I", IEEE Transactions on Communications, pp. 942-956, September 1975

D. L. Fadley, "Multi-Gigabit Laser Communications Electronics", unpublished manuscript, Motorola, Inc.

B. H. Batson and G. K. Huth, "Convolutional Coding at 50 Mbps for Shuttle Ku-Band Return Link", Proceedings of the International Telemetry Conference, Los Angeles, California, pp. 175-181, September 28-30, 1976

"Model 720-02 Bit Synchronizer", EMR Telemetry, Sarasota, Florida

"FR-3000 Multiband Instrumentation Recorder", D249-7-75, Ampex Corporation, Data Products Division, Redwood City, California

"LSI Data Communication Device Technology Available for Commercial and Industrial Applications", Computer Design, pp. 98-101, December 1976

J. P. Harris et al., "The IBM 3850 Mass Storage System: Design Aspects", Proceedings of the IEEE, Volume 63, No. 8

"Telemetry Standards", IRIG 106-75

E. P. Greene, "Telemetry Packetization for Improved Mission Operations", Proceedings of the International Telemetry Conference, Los Angeles, California, pp. 337-344, September 28-30, 1976

Performance Specification for Services Via The Tracking and Data Relay Satellite System", S-805-1, Goddard Space Flight Center, Maryland, November 1976

"d/pad One Data Processor, Analyzer, and Display System", IED, San Diego, California

J. A. Heller and I. M. Jacobs, "Viterbi Decoding for Satellite and Space Communication", IEEE Transactions on Communications, pp. 835-848, October 1971

J. S. Gray, "120 Mb/s and 240 Mb/s Bit Synchronizer-Signal Conditioners for NASA High Data Rate Applications", Proceedings of the International Telemetry Conference, Los Angeles, California, pp. 358-368, September 28-30, 1976

"Error Control Products", Linkabit Corporation, San Diego, California

"Model 711 PCM Decommutator", EMR Telemetry, Sarasota, Florida

J. Mavor, "Signal Processing Applications of Charge-Coupled Devices", The Radio and Electronic Engineer, pp. 412-420, August/September 1976

R. O. Leighou, "High-Density Digital Data Recording/Reproducing System", Proceedings of the International Telemetry Conference, Los Angeles, California, pp. 496-505, September 28-30, 1976

"Tracking and Data Relay Satellite System (TDRSS) Users' Guide, STDN 101.2", Goddard Space Flight Center, Maryland, January 1978

W. W. Wu, "New Convolutional Codes - Part III", IEEE Transactions on Communications, pp. 946-955, September 1976

"Magnetic Tape Recorder Selection Guide", Bell & Howell, Datatape Division, Pasadena, California

J. E. Goodwin, "A Digital Data Interleaver", Proceedings of the International Telemetry Conference, Los Angeles, California, pp. 514-521, September 28-30, 1976

7. GROUND PROCESSING ELEMENTS

Ground processing elements represent a significant portion of the NASA end-to-end data system in terms of both cost and performance. Processing elements discussed in this report include hardware, software, and firmware, as depicted in Figure 7-1. This section presents the state of the art, trends, and projected developments through 1985 in processers (micros, minis, large-scale, and super-scale), storage devices (main memories and auxiliary memories), and firmware devices. Software (languages, operating systems, and methodology) is discussed in Section 11. Certain data storage elements (e.g., archival data storage devices) are discussed in Section 8. Communication elements are discussed in Section 9, and information presentation elements (displays, pointers, and platters) are covered in Section 10. Data base management systems are discussed in Section 8. Space processing hardware elements are discussed in Section 2. Space data storage and space data handling are covered in Sections 3 and 4, respectively. Pre-processing elements, which overlap with ground processing elements, are presented in Section 6

Figures 7-2 through 7-4 present the different hardware elements in terms of their Level 2 and higher elements. Subsequent sections present the state of the art, trends, and projected developments for the different levels.

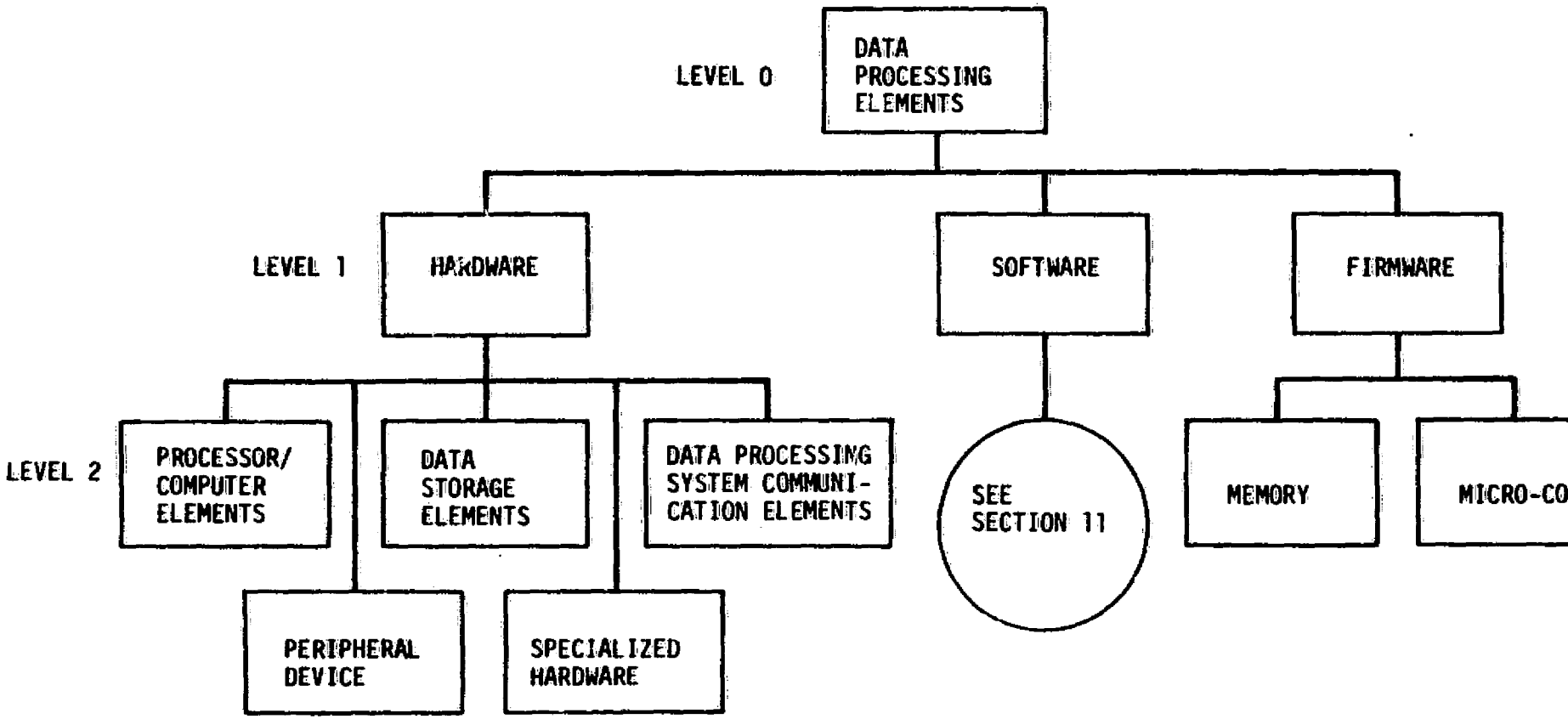


FIGURE 7-1. DATA PROCESSING ELEMENTS

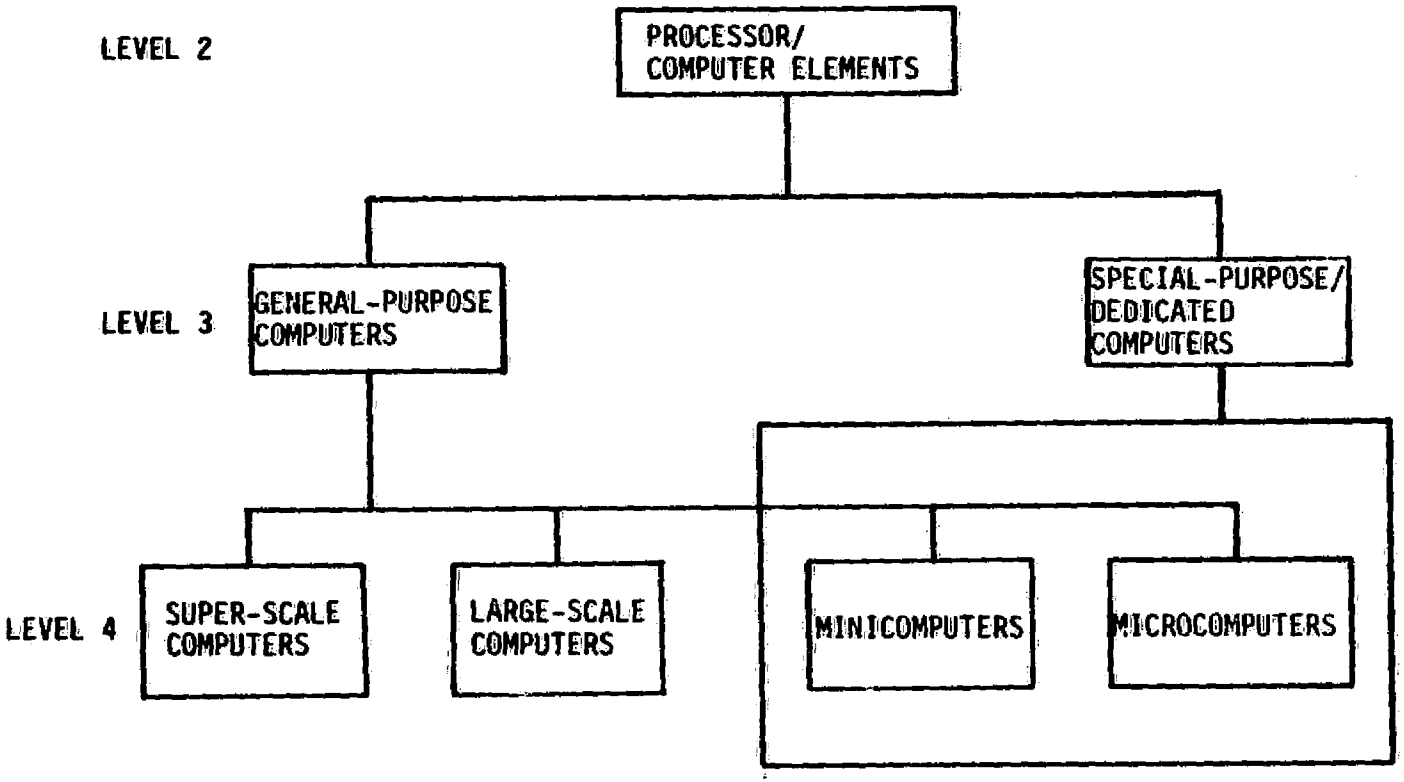


FIGURE 7-2. PROCESSOR/COMPUTER ELEMENTS

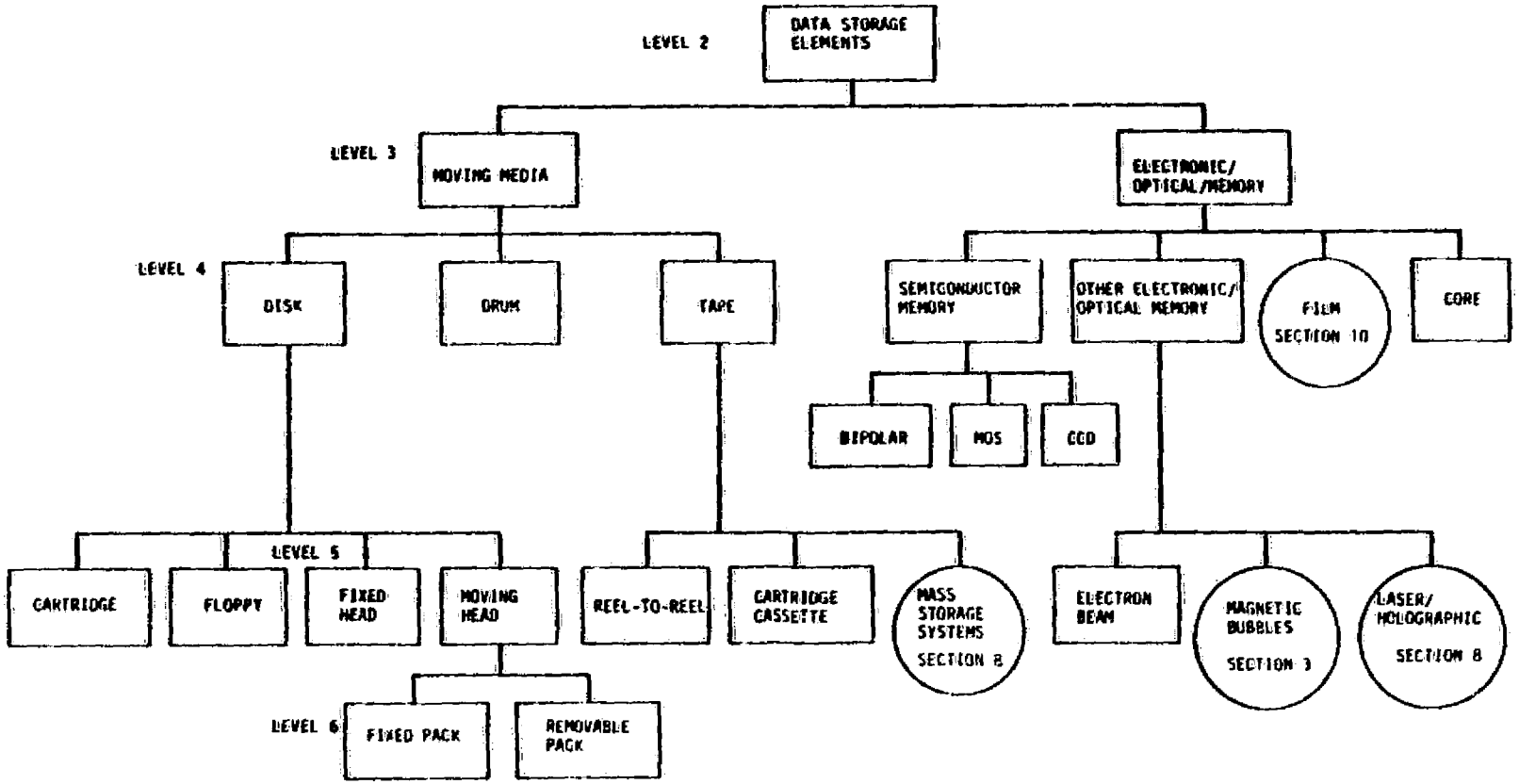


FIGURE 7-3. DATA STORAGE ELEMENTS

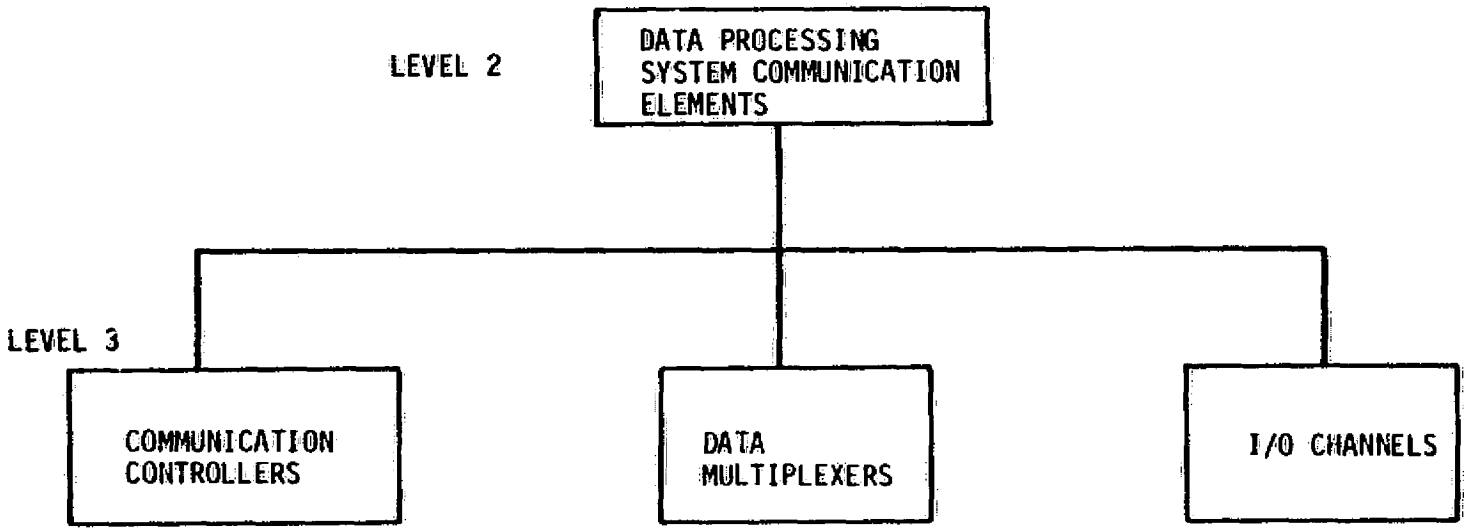


FIGURE 7-4. DATA PROCESSING SYSTEM COMMUNICATION ELEMENTS

7.1 PROCESSOR/COMPUTER ELEMENTS

Processor/computer elements are presented in terms of four areas of processing capability that are considered to be representative of the processing technology in general. The four areas, illustrated in Figure 7-2, are microcomputers, minicomputers, large-scale computers (defined herein to be the upper end of the mainframe manufacturer's line), and super-scale computers. Precise distinctions between each of these classes of computers are difficult to establish and are becoming more difficult to define daily. These boundaries, as defined in the various subsections, are illustrated in Figure 7.1-1 for each of the four areas of processing capability. Figure 7.1-1 also includes projections for the throughput performance of single-chip microcomputers and super-scale computers for the 1985 timeframe.

One notable feature of this figure is the tremendous increase shown in processing power projected by 1985 for single-chip microcomputers. This growth, as discussed in this section and Section 1, will significantly impact the architecture of large-scale computer systems by 1985. In general, large-scale computers of 1985 will distribute functions now handled by the central processor to both remote peripherial processors and to local peripheral devices within the central computer complex.

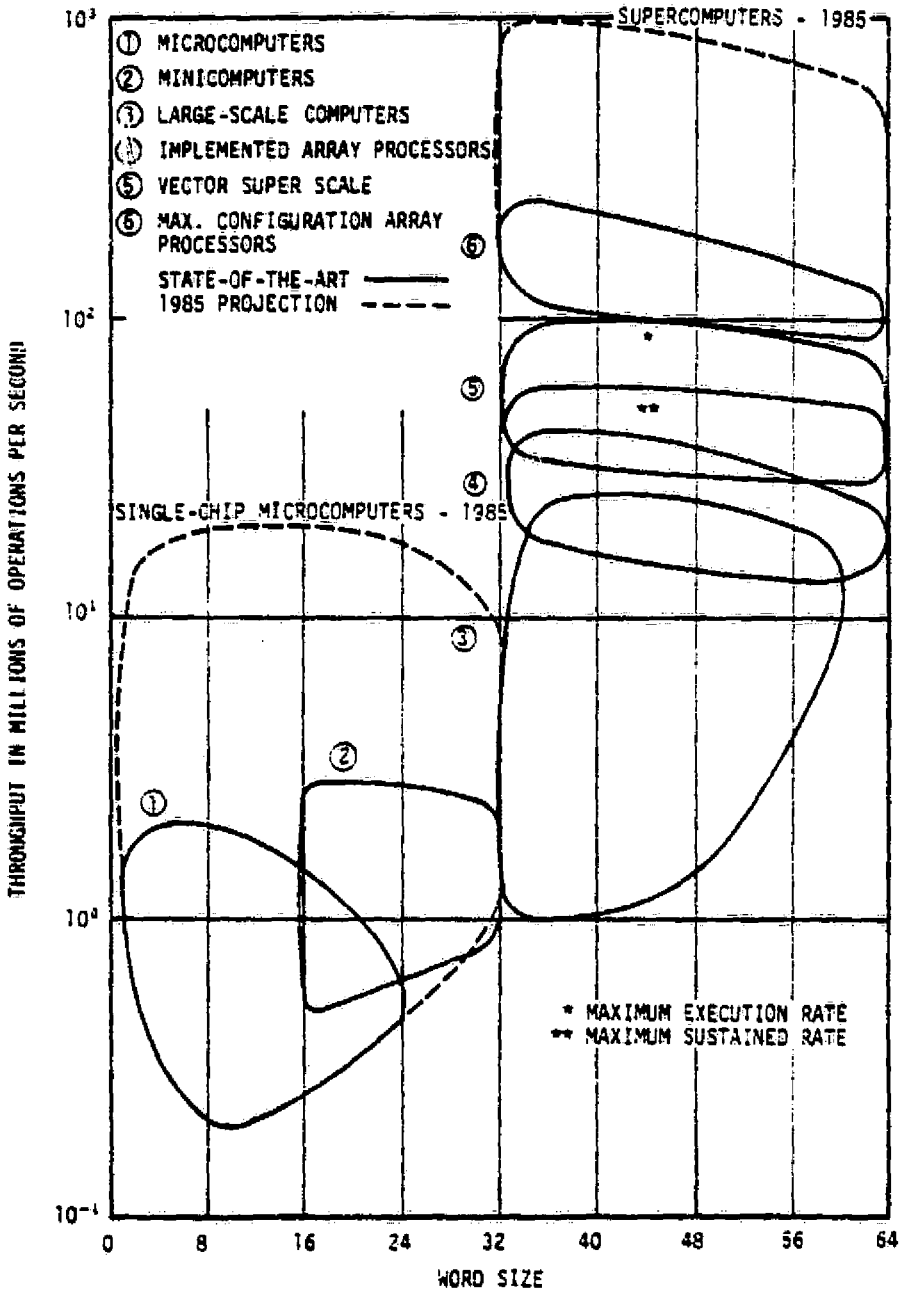


FIGURE 7.1-1. PERFORMANCE CHARACTERISTICS FOR EACH OF THE FOUR DEFINED AREAS OF COMPUTERS

7.1.1 Microcomputers

To identify the characteristics that distinguish a microcomputer from any other computer, it is convenient to compare it with its predecessor, the minicomputer. Both the microcomputer and the minicomputer acquired their names because of relative size differences when comparing them with their predecessors. Like the minicomputer, the most obvious microcomputer distinctions at the time of first introduction were in physical size, cost, word size, limited processing capability, and input/output (I/O) mechanism complexity. However, as time passes the distinctions between the microcomputer and the minicomputer become less and less clear, especially in the capability borderline region (the low end of the minis and the high end of the micros).

Microcomputers could be identified initially by their exclusive use of large-scale integration (LSI) circuitry, whereas minicomputers initially used either discrete components or small-scale integrated circuits. Today, most minicomputers use LSI and many microcomputers use a multichip CPU (bit slice) resembling those used in many minicomputers. Microcomputers were initially constrained by memory size, whereas minicomputers of the same time period were not. Today, some microcomputers can use memories to 1 Mbyte (not directly addressable). In the not-too-distant past, high-level languages such as FORTRAN and BASIC were only available on minicomputers, but now most microcomputers have one or both of these. The software packages available for minicomputers are still much more complex than those available for microcomputers. In several instances, microcomputer throughput equals or exceeds that of some minicomputers. However, for this report, the distinguishing features between the microcomputer and the minicomputer are constrained to those presented in Table 7.1.1-1.

TABLE 7.1.1-1. COMPARATIVE CHARACTERISTICS OF CURRENT MICROCOMPUTERS VERSUS MINICOMPUTERS

CHARACTERISTICS	MICROCOMPUTERS	MINICOMPUTERS
Cost	$500 to $5,000	$2,000 to $100,000
Size	Desk top	19-in. rack mounted
Software	Being developed, relatively simple	Already established, fairly complex
Input/Output	Front panel or control terminal interface built in; serial printer, cassette tape floppy disk, etc., interface optional	Front panel and/or control terminal interface built in; line printer, reel-to-reel tape, hard disk, etc., interface optional

7.1.1.1 State-of-the-Art in Microcomputers - The term "microcomputer" is currently used to describe any of a large variety of computer types constructed from high-complexity integrated circuit building blocks. These building blocks are derived from a number of different large-scale integrated bipolar and MOS circuit technologies that include TTL, Schottky TTL, I^2L, PMOS, NMOS, CMOS, and CMOS/SOS.

In general, the bipolar building blocks are characterized by higher-speed operation and lower complexity than the MOS microcomputer building blocks. For example, state-of-the-art Schottky TTL 4-bit-slice processor elements contain about 300 gates (limited by power dissipation) and operate at a clock frequency of approximately 10 MHz. Depending on the circuit family selected, anywhere from 5 to 40 LSI TTL building blocks are required to implement a typical 16-bit microcomputer. The microcomputer designs that contain only a few TTL LSI devices usually execute a fixed instruction set, whereas the more complex TTL microcomputers utilize a microprogrammed architecture that allows the microcomputer to emulate various instruction sets simply by changes in the microcode. Emulation of previous computer architectures makes it possible to use existing applications software, thus providing significant savings in software development costs.

The MOS microcomputer building blocks are generally much more complex than the bipolar circuits just discussed. For example, Intel Corporation recently reported the development of a dual-processor, microprogrammable, N-channel MOS chip that contains approximately 22,000 transistors. Because of the lower power consumption per gate and high circuit density of the MOS-type microprocessor building blocks, it is possible to build a single-chip microcomputer. These single-chip microcomputers are currently finding widespread application in a variety of consumer product applications. Programming these single-chip designs is either accomplished during the fabrication process or by using erasable programmable read-only memories (EPROMs) that are contained on the chip. Multi-chip microcomputer designs are currently

available with single-chip microprocessors providing up to 16-bit parallel operation. State-of-the-art 16-bit microprocessors feature direct addressing of up to 64 Kbytes of memory, multiple addressing modes, and average instruction execution speeds of approximately 1.0 μsec at a power dissipation of 0.5 W.

Another bipolar technology that provides both high circuit complexity and low power dissipation is integrated injection logic (I^2L). Although I^2L is somewhat slower than TTL, it has a power-delay product less than 1 pJ (TTL typically exceeds 10 pJ). The circuit packing density of I^2L is much greater than TTL because the circuit is configured such that the isolation islands used in TTL between transistors are not required. Microcomputer building blocks that are currently available in I^2L technology range in complexity from 4-bit-slice processor elements that typically contain 1,600 gates up to complete 16-bit microprocessors.

Economically, the emergence of the microcomputer is most important because it has provided the system designer with the basic elements necessary to develop relatively complex solutions to specialized processing and control hardware design problems at low cost. An additional benefit provided by microcomputers is that they have provided an increase in commonality of usage of integrated circuits among diverse applications, thereby providing a reduction in the proliferation of special-purpose, high-cost circuit types.

The architecture of today's microcomputer, though assembled with a particular manufacturer's unique enhancements, consists of four basic elements: an arithmetic and logic unit (ALU), a control unit, a storage unit with program and data store areas, and an input/output (I/O) unit (see Figure 7.1.1.1-1). The combination of an ALU, control unit, and registers (used in arithmetic and logical operations to hold data, memory addresses, etc.) is called a Central Processing Unit (CPU).

The normal architecture for a multi-chip set is to have the CPU on one chip and the storage and I/O on separate chips. The limitations of this configuration are that the data width (4, 8, 16 bits, etc.) is fixed and the instruction set is located within the CPU and is not changeable. The bit-slice bipolar microcomputers use multiple chips by design to implement the CPU. Hence the data width can be tailored

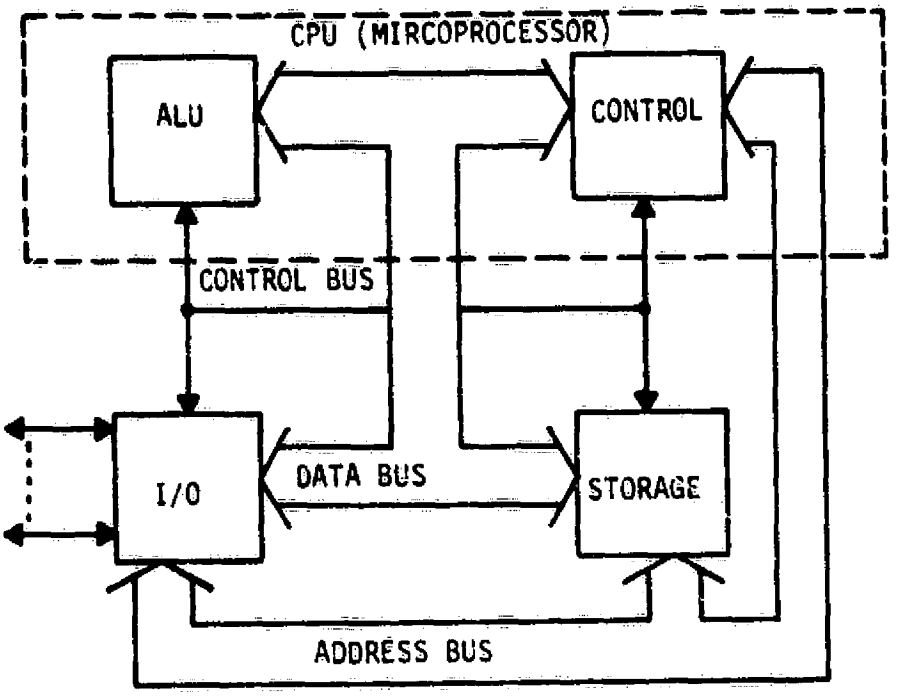


FIGURE 7.1.1.1-1. TYPICAL MICROCOMPUTER FUNCTIONAL BLOCK DIAGRAM

as required to meet the designer's requirements (4, 8, 16...32, 48, 64 bits). Most bipolar bit-slice microcomputers use a 4-bit slice as their basic building block. The maximum data width size for state-of-the art MOS microcomputers is 24 bits (Four Phase Systems System IV), with most designs providing 16 bits or less.

Microcomputer acceptance and utilization throughout the commercial, military, and aerospace industries continues to increase exponentially. Table 7.1.1.1-1 shows some of the more typical uses of microcomputers at the present time. A summary of the characteristics of state-of-the-art microcomputers is presented in Table 7.1.1.1-2. The cost of a microcomputer system can range anywhere from under $500 to well over $50,000. The high-cost end is determined by the types and numbers of peripherals, types and sizes of memory/storage, the number of microcomputers utilized in the system configuration, and the capabilities of the software that either comes with the system or must be developed. Note that in many instances the cost of the peripheral devices can be far greater than the cost of the microcomputer and its memory. Those systems costing about $500 may not have any peripherals. The consumer products and automative industries are two of many users of low-cost microcomputers.

TABLE 7.1.1.1-1. APPLICATIONS AND CAPABILITIES OF MICROCOMPUTERS AS A FUNCTION OF WORD SIZE AND CONFIGURATION

	CONTROL		DATA PROCESSING	
	DEDICATED 4-bit	GENERAL PURPOSE 8- to 16-bit	DISTRIBUTED 8-bit	CENTRAL 16-bit
	Single chip High volume (10,000 piece lots) Very low cost ($3)	Single chip Medium volume Low cost	Multi-chip Low volume Medium cost	Multi-chip Very low volume High cost
APPLICATION	Consumer oriented: TV tuner Nonvideo games Appliances Entertainment	Industrial: Instrumentation Automotive Peripheral controller Machine controller	Business and real-time control: Intelligent terminals Industrial control Process control	Real-time data processing: Data base Data processing Big business Scientific
CAPACITY	Limited chip capability: 4-bit data manipulation Small program ROM Small on-chip RAM Limited I/O Not expandable	Fairly extensive chip capacity: 8- or 16-bit data handling 2 Kbyte program ROM 128 bytes of RAM Extensive I/O Expandable	Extensive chip capacity: 8- or 16-bit data handling 64 Kbytes of ROM 64 Kbytes of RAM Extensive I/O Unlimited expansion	Very extensive chip capacity: 16- to 32-bit data handling 64 Kbytes of ROM 256 Kbytes of RAM Unlimited expansion

TABLE 7.1.1.1-2. STATE OF THE ART IN MICROCOMPUTERS

CHARACTERISTICS	RANGE		TYPICAL
CPU (microprocessor)	Choice from over 50 vendors		
Data Word Length (bits)	4	64	8, 16
Instruction Word Length (bits)	4	64	8, 16, 24
Type of Logic Circuitry	NMOS, BIPOLAR, PMOS, CMOS		NMOS
Cycle Time (μsec)	0.125	13	1
Add Time (μsec)	0.2	20	2
Number of Instructions	32	197	78, 72
Total No. of Registers	2	104	7, 6, 8
MIPS	0.1	4	1.0
Random-Access Memory			
Type	NMOS, CMOS, BIPOLAR, PMOS		NMOS
Cycle Time (nsec)	100	2,000	450, 500
Capacity (bytes)	1K	128K	64K
Word Length (bits)	4	16	8, 16
Read-Only Memory			
Type	NMOS, PMOS, CMOS, BIPOLAR		NMOS
Cycle Time (nsec)	65	1,500	450
Capacity (bytes)	256	512K	64K
Word Length (bits)	8	16	8
Programmable ROM			
Type	NMOS, BIPOLAR, PMOS, CMOS		NMOS
Cycle Time (nsec)	65	1,600	450
Capacity (bytes)	256	128K	64K
Word Length (bits)	8	16	8
Input/Output Control			
I/O Word Length (bits)	1	20	8
No. of I/O Channels	2	512	256
No. of Interrupts	1	256	8
Maximum I/O Rate (words/sec)	13K	4M	1M, 2M
Software			
Assemblers	Resident and/or cross assemblers available		Assembler
High-Level Languages	PL/M, BASIC, FORTRAN - FOCAL, PLUS, COBOL (most available as cross compilers only)		BASIC (resident), FORTRAN
Simulators	May be available		Yes
Monitor/Operating Systems	Monitor usually available		Small (1 Kbyte) monitor
Physical Characteristics Temperature (°C)	0 to 50	-55 to 125	0 to 70
Cost of Basic System	$500	$14,500	$1,500
MTBF	NA	NA	NA

7.1.1.2 Trends in Microcomputers - Since the introduction of the first chip microprocessor (the CPU of a microcomputer) in 1971, there has been a continual explosion in the advancement of microcomputers. This explosion, however, is not singular and appears to be branching into three specific areas: 1) commercial microprocessors in consumer products, 2) commercial microcomputers functioning as general-purpose or special/dedicated-purpose machines in both real-time and non-real-time environments, and 3) military/aerospace ruggedized microcomputers. Technology advancements in the near future will be primarily directed toward lowering production costs and further market penetration of the extremely large volume consumer product markets at the low end of the microcomputer scale.

Other trends in microcomputers include continuing development of more complex chips, improvements in operating speeds, and lower costs per unit of performance. As explained in Section 1, the number of components or circuit elements contained in the most complex integrated circuit chips should continue to double each year through 1985. As the bridge from optical fabrication techniques to suboptical (or submicron) techniques is crossed around 1980, there may be a temporary slowdown in the circuit complexity growth rate. However, once the manufacture of suboptical devices becomes widespread, the doubling in complexity each year should resume and continue for at least another 10 years.

The result of higher chip complexity will be lower cost both per unit of complexity and per unit of performance. Figure 7.1.1.2-1 illustrates how the cost per component (resistor, transistor, etc.) should decrease as a function of time through 1985. For example, a 16-bit microcomputer chip with a 4-Kbyte ROM and a 4-Kbyte RAM will need approximately 10^5 circuit elements and should sell for about $10 in 1980 (constant 1977 dollars). By 1985, 32-bit microcomputer chips with larger on-chip memories will be available for about the same price.

The speed of LSI/VLSI microcomputers will increase by more than an order of magnitude by 1985 as a result of smaller-geometry devices

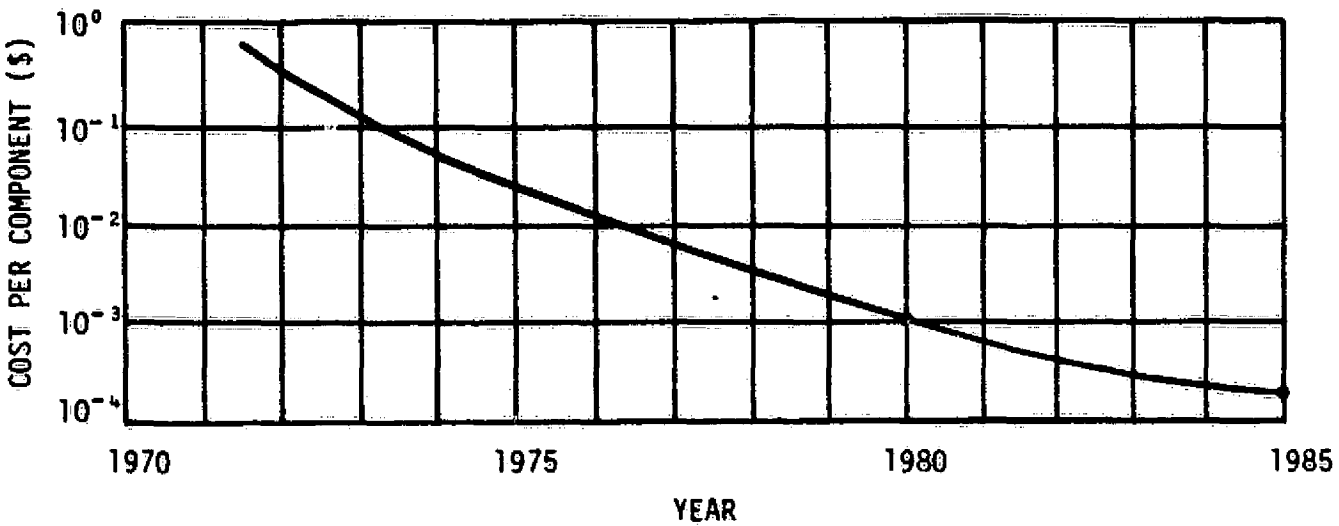


FIGURE 7.1.1.2-1. TRENDS IN MICROCOMPUTER COMPONENT COST

and the development new high-speed, low-power circuit technologies. Typical gate delay times for LSI microcomputer elements employing NMOS and I^2L technologies are currently about 10 nsec. By 1985, typical gate delay times for VLSI microcomputer circuits will be less than 0.4 nsec, or more than 25 times faster than current LSI technology. This is illustrated in Figure 7.1.1.2-2.

Some of the newer logic technologies currently under development that have demonstrated both high speed and low power consumption and thus show promise for implementing next-generation microcomputers include; DMOS, VMOS, elevated electrode logic, static induction transistor logic, E-mode GaAs MESFET, Schottky GaAs MESFET, CCD logic, and Josephson junctions. Most of these technologies have power-delay products that range from 20 to 1,000 times less than high-density ECL and can be faster than ECL.

One conclusion that can be made concerning 1985 computer technology is that by that time all production computers will employ some form of VLSI microcomputer technology. Another important conclusion

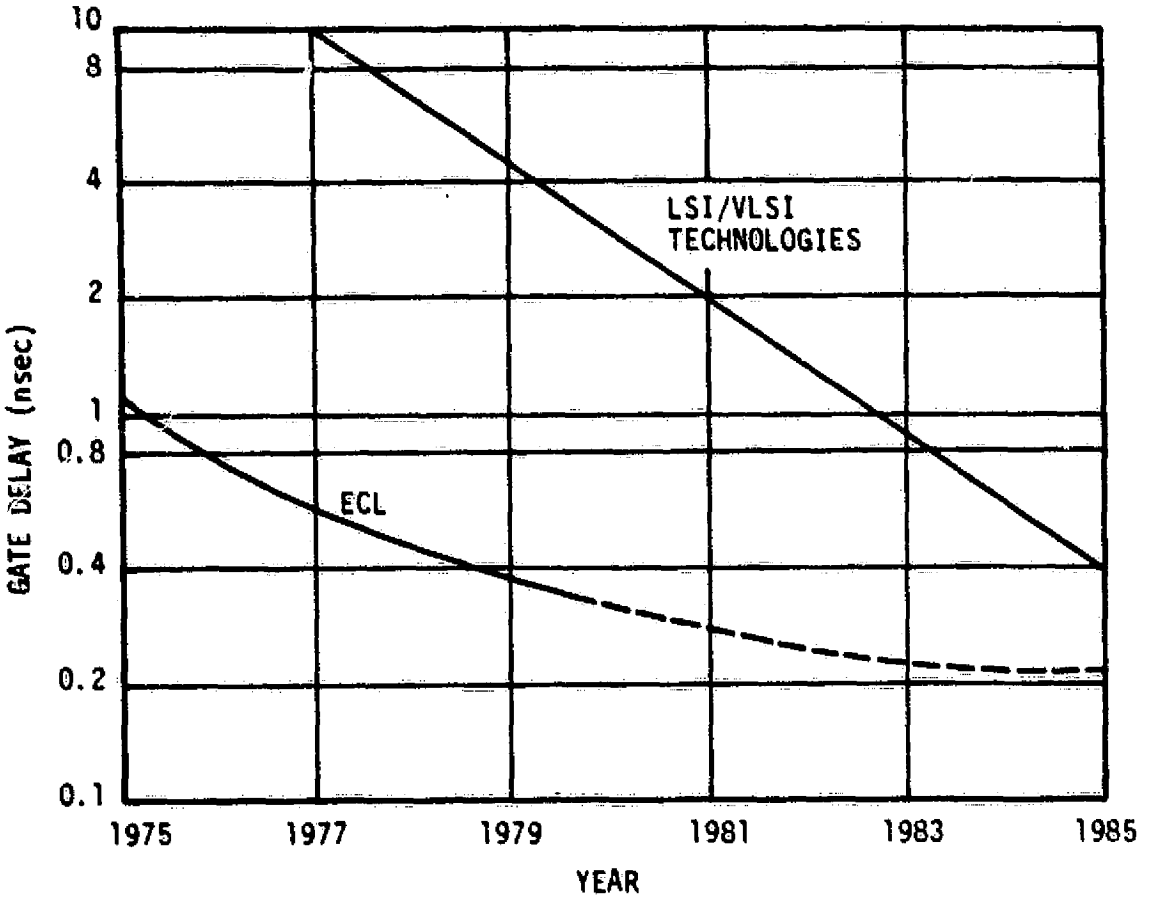


FIGURE 7.1.1.2-2. TRENDS IN MOS LSI MICROCOMPUTER PROCESSING SPEED

is that, from now into the mid-1980's, the trend in the time span from initial computer hardware introduction to hardware obsolescence can be expected to decrease from its current value of about 4 to 5 years to less than 3 years. This shortened time period, coupled with the manufacturers' need for higher dollar investments in advanced development and production equipment, will cause a large percentage of the existing manufacturers to leave the field.

7.1.1.3 Projected Developments in Microcomputers - By 1980, single-chip microcomputers with 32-bit word lengths will be available with performance speeds of 2 million instructions per second (MIPS) or more. Using a multiple chip microcomputer configuration, throughput performance of 5 MIPS is expected by 1980, and by 1985 approximately 20 MIPS should be possible. Coupled with speed enhancements will be the emergence of many microcomputers with specialized architectures, some of which will use many of the performance techniques utilized in the large computers of today. An example might be a microcomputer containing a pipeline for matrix operations or a microcomputer especially suited to a Fast Fourier Transform. Space and power requirements for comparable systems will drop to about 1/50th of today's requirements. System reliability will be substantially higher due to both the reduction in the number of circuit interconnections and the opportunities that a disassociated system architecture provides for the introduction of system element replacement and redundancy.

A major problem foreseen with the microcomputer is the need to assemble federated or democratic networks of computers to achieve the high throughput levels required by the user. The system software control and data base management mechanisms for accomplishing this networking are not well known and are only in the preliminary stage of investigation. It is anticipated that by 1985 these problems will be partially resolved and that development of the software controls and data base management will in effect be automated to the point where some degree of automatic software generation will be available.

Table 7.1.1.3-1 contains projections of the state of the art in microcomputers for the 1980 and 1985 timeframes.

In the military/aerospace area, utilization of improved VLSI microcomputers will proliferate. Savings in size, weight, power consumption, and cost in addition to enhanced system life and improved reliability will be the driving factors. By 1985, the military/aerospace community will be utilizing their systems approach to develop

TABLE 7.1.1.3-1. PROJECTED DEVELOPMENTS FOR TYPICAL MICROCOMPUTERS

CHARACTERISTICS	1980	1985
CPU (microprocessor)	Choice from 30 vendors	Choice from 20 vendors
Data Word Length (bits)	16, 32	32, 64
Instruction Word Length (bits)	8, 16, 32	16, 32
Type of Logic Circuitry	NMOS, VMOS, HMOS, DMOS	VMOS, HMOS, DMOS, GaAs
Cycle Time (usec)	0.1	0.025
Add Time (usec)	0.2	0.05
Number of Instructions	100 (1/2 selectable)	150 (all selectable)
Total No. of Registers	16, 32	16, 32
MIPS	5	20
Random-Access Memory		
Type	NMOS, VMOS, HMOS, DMOS	VMOS, HMOS, DMOS
Cycle Time (nsec)	300	300
Word Length (bits)	16, 32	32, 64
Capacity (bytes)	64K	256K
Read-Only Memory		
Type	VMOS, HMOS, DMOS I^3L	VMOS, HMOS, DMOS
Cycle Time (nsec)	200	200
Word Length (bits)	16, 32	32, 64
Capacity (bytes)	64K	64K
Programmable ROM		
Type	VMOS, HMOS, DMOS	VMOS, HMOS, DMOS
Cycle Time (nsec)	200	200
Word Length (bits)	16, 32	32, 64
Capacity (bytes)	64K	64K
Input/Output Control		
I/O Word Length (bits)	8, 16	16, 32
No. of I/O Channels	Selectable	Selectable
No. of Interrupts	Selectable	Selectable
Maximum I/O Rate (words/sec)	1M	2M
Software		
Assemblers	Assembler	Assembler
High-Level Languages	BASIC, FORTRAN (available in resident and cross-compiler versions)	BASIC, FORTRAN Higher Languages (resident)
Simulators	Yes	Yes
Monitors/Operating	Small-sized monitor (a few Kbytes)	Medium-sized monitors (several Kbytes)
Physical Characteristics Temperature (°C)	0 to 70, some -55 to 125	0 to 70, some -55 to 125
Cost of Basic System	$1,000	$500
MTBF	NA	NA

the software first and then assemble mission hardware to fit the software. This approach will eliminate current problems experienced with obsolete hardware caused by early selection and long development/ production time schedules. System spares will be procured with the production hardware. This will eliminate a future problem of no manufacturer available to fabricate obsolete logic. Since hardware cost will have fallen to low dollar levels, the cost of excess purchases of spares will not be significant. Software will still be the driving cost factor with the systems through 1985. The system operational life of 10 to 20 years will not change. System upgrades will be accomplished with state-of-the-art hardware assembled to emulate the desired architecture. Advancements in radiation hardness tolerance levels will occur as an unexpected byproduct in commercial developments (no extra cost) and as a direct result of continued funded research and development. In addition, system designers will pay closer attention to the computer operating environment and utilize commercially acceptable items where applicable.

7.1.2 Minicomputers

7.1.2.1 State of the Art in Minicomputers - The capabilities of today's minicomputers cover a very broad spectrum in terms of the size and speed of CPU and memory, software capabilities, instruction sets, and related areas. In view of the wide range of capabilities offered, the summary of the basic characteristics of state-of-the-art minicomputers presented in Table 7.1.2.1-1 includes one listing for the average or typical minicomputer and one listing for the typical super-minicomputer. The cost of a minicomputer system can range anywhere from under $1,000 to well over $200,000, with computing power increasing rapidly with increasing cost. Many of the super-minicomputers take advantage of high-speed bipolar integrated logic circuits, bipolar cache memories, and/or memory interleaving to achieve high-throughput performance levels. The typical minicomputer uses NMOS technology, sacrificing some speed for lower cost.

Both core and MOS main memories are found in minicomputer systems, with neither technology dominating the other at the present time. Faster bipolar main memories are available on some systems as a customer-selected option, but the cost differential is too great for widespread use. As an enhancement to system performance and throughput, a few super-minicomputers offer virtual memory capabilities and several super-minicomputers offer bipolar cache memories and/or memory interleaving for the purpose of increasing processing speed and decreasing average memory cycle time. The use of these techniques is particularly significant for those systems still using the slower core main memory. Examples of the effectiveness of these techniques are:

- The 240-nsec bipolar cache memory on the PDP-11/70 reduces the 980-nsec core memory cycle time to an average of 400 nsec.
- Memory interleaving and dual-instruction look-ahead on the Interdata 8/32 reduce the 750-nsec core memory cycle time to an average of 300 nsec.

TABLE 7.1.2.1-1. STATE OF THE ART IN MINICOMPUTERS

CHARACTERISTICS	TYPICAL MINICOMPUTER	SUPER-MINICOMPUTER
CPU		
Word Length (bits)	16	16 or 32
Type of Logic Circuitry	NMOS	TTL
Cycle Time (μsec)	0.250 to 0.400	0.160 to 0.300
Add Time (μsec)	1 to 3	0.3 to 0.7
Number of Directly Addressable Words	256 to 32K	32K to 1,024K
Number of Instructions	100 to 150	120 to 300
Total Number of Registers	0 (for some stack machines) to 32	0 to 32
MIPS	0.25 to 0.60	0.7 to 2.0
Main Memory		
Type	Core or NMOS	Core or NMOS
Cycle Time (μsec)	0.8 to 1.2; 0.5 to 0.7	0.5 to 0.98; 0.2 to 0.7
Minimum Capacity (words)	4K to 32K	16K to 64K
Maximum Capacity (words)	64K to 256K	128K to 4M
Word Length (bits)	16 plus parity for some systems	16 or 32 plus error correction for some systems
I/O Maximum Rate (words/sec)	1M to 2M	1M to 6.67M

TABLE 7.1.2.1-1 - Concluded

CHARACTERISTICS	TYPICAL MINICOMPUTER	SUPER-MINICOMPUTER
Software		
Assemblers	Assembler	Assembler, macroassembler
High-Level Languages	FORTRAN, BASIC	FORTRAN, BASIC, COBOL, RPG
Operating Systems	Batch, real-time	Batch, real-time, time-sharing, interactive
Physical Characteristics		
Volume (in^3)	4,500	50,000
Voltage (Vac)	115 at 60 Hz or 220 at 50 Hz	115 at 60 Hz or 220 at 50 Hz
Operating Temperature Range (°C)	0 to 50	0 to 50
Humidity Tolerance (%)	0 to 95	0 to 95
Cost of Basic System (CPU, power supply, front panel, and minimum memory in a chassis)	$5,000 to $20,000	$20,000 to $100,000
MTBF (hr)	NA	NA
Comments		Many newer systems use bipolar cache memories and/or memory interleaving

System reliability has been enhanced through the use of memory parity checking, storage protection, and LSI logic circuits. Use of LSI circuits in minicomputers has permitted the manufacturers to make the CPUs smaller, faster, and inherently more reliable. Parity checking and storage protection features, if not included as part of the basic system, are offered as options to the customer. As an additional feature, most of the recently announced minicomputers are including a 1-bit error correct and a 2-bit error detect capability.

The instruction sets available on minicomputers vary greatly, both in number of instructions and complexity. Most of the super-minicomputer instruction sets now include hardware multiply/divide and hardware floating point arithmetic, either as optional or standard equipment. Availability of control storage on super-minicomputer systems in the form of ROM, PROM, or WCS (writable control storage) is also widespread. This allows the vendor and/or user to tailor the mini-computer's internal processing capabilities to better meet the users' needs. User-accessible microprogrammability through WCS greatly increases the flexibility of today's systems.

The architectures of today's minicomputers fall into three classes based on the type of bus structure incorporated into the system design. These classes, single bus, compatible input/output bus and multiple bus structures, are illustrated in Figures 7.1.2.1-1 through 7.1.2.1-3 (Ref. 7-1).

Single bus structures, typified by the patented Unibus of DEC's PDP-11 family of minicomputers, use one bus that interconnects all system elements (processor, memory modules, and input/output (I/O) controllers). Protocol for use of the bus is identical for all elements, thus making it possible for various elements to communicate directly without processor (CPU) control, as an input/output device can communicate directly with memory or another I/O device. Note, however, that while such as operation is taking place, the CPU cannot use the bus and thus remains idle unless it is busy with a task that does not require bus access (e.g., instruction decoding, execution of a register to register add).

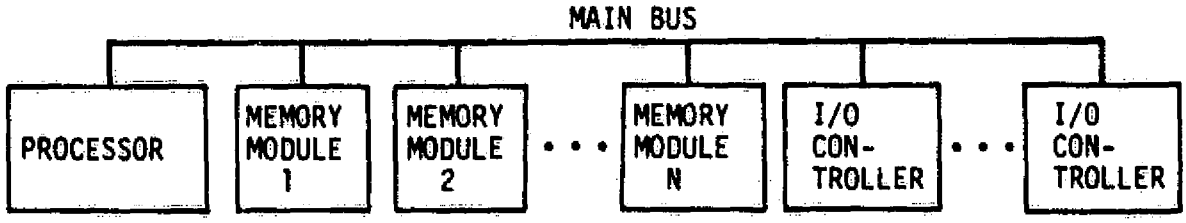


FIGURE 7.1.2.1-1. MINICOMPUTER SINGLE BUS ARCHITECTURE

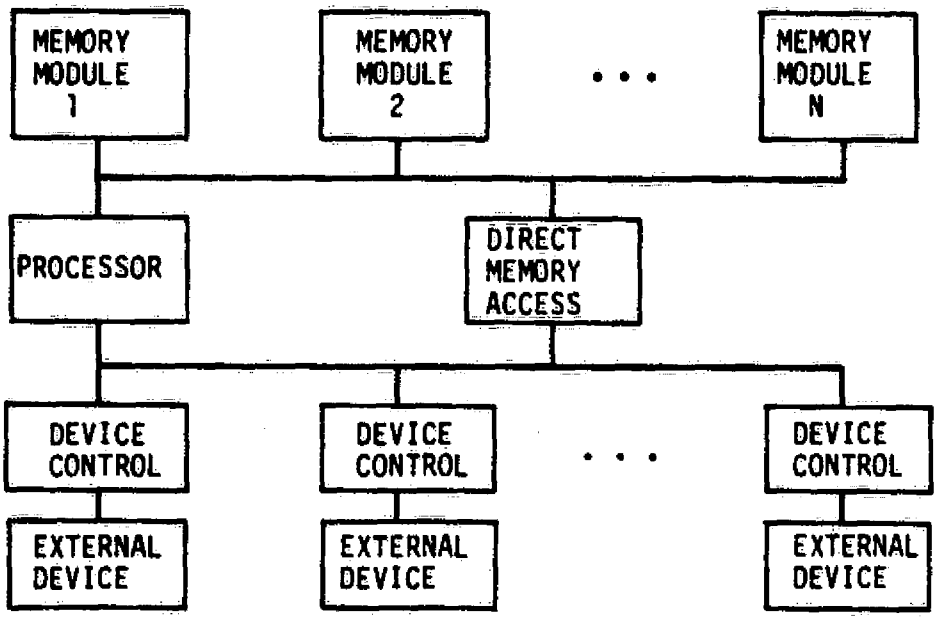


FIGURE 7.1.2.1-2. MINICOMPUTER COMPATIBLE INPUT/OUTPUT BUS ARCHITECTURE

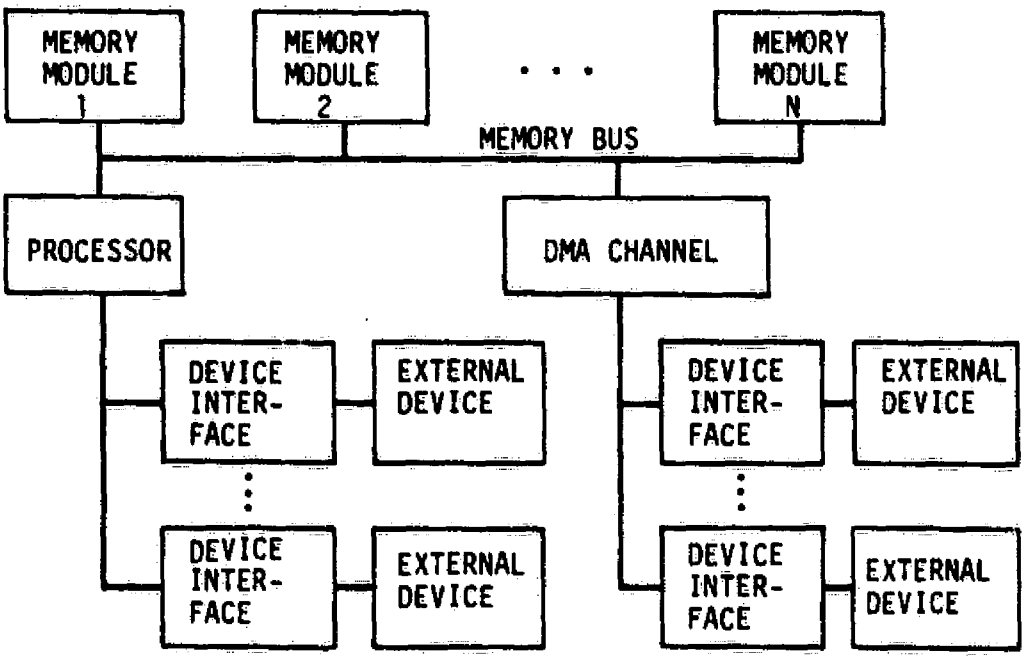


FIGURE 7.1.2.1-3. MINICOMPUTER MULTIPLE BUS ARCHITECTURE

Compatible input/output bus structures require two buses that use compatible protocols. One bus carries communications only between the CPU and the input/output controllers, the other only between the CPU and the main memory modules. The need for compatible protocols arises when the Direct Memory Access (DMA) unit is used to provide a path for data flow between an I/O controller and a memory module. The speeds of the two buses do not have to be equal, thus permitting slower and more economical I/O controllers to be used.

Multiple bus structures further divide the I/O bus into a programmed input/output bus and a DMA input/output bus. (This structure is used by a large number of minicomputer manufacturers, including Data General, NCR, Texas Instruments, and Varian.) Unlike the compatible I/O bus structure, each element is assigned to a particular I/O bus and cannot be used for both buses -- the programmed I/O and DMA I/O. The speed and protocol of each bus can be tailored to the specific requirements of the users' operation. As a result, the bus DMA may be faster or slower than on a compatible I/O bus structure.

Section 4 discusses the technology for implementing data buses, with emphasis on fiber-optic devices, which will see wide use in future data bus designs.

7.1.2.2 Trends in Minicomputers - Since their introduction approximately a decade ago, minicomputer throughput capabilities have increased steadily while their costs per processed instruction have decreased. Technological and manufacturing innovations will continue these trends over the next several years, as illustrated graphically in Figures 7.1.2.2-1 through 7.1.2.2-5. Super-minicomputers will rival the far more costly medium-scale computers in terms of processing power and flexibility. Increased use of microprogrammed logic will enhance the flexibility of minicomputers in general. As a result of these trends, there will be an ever-widening market for these small, economical, and surprisingly fast computers. Although most minicomputers are now used in process control and laboratory instrumentation, increasing emphasis is being placed on their use in distributed processing and conventional business data processing applications. Primarily because of the lack of sophisticated software, minicomputers are really just beginning to make a significant impact in the business world. There is also a definite movement by minicomputer manufacturers toward the development of complete systems, with the computer mainframe, peripherals, and sophisticated software all provided and working together to totally meet user needs effectively (turn-key systems). This movement will become attractive to the potentially very large market of business data processing users.

There is a clear-cut industry trend toward increased use of LSI logic circuitry in minicomputers. NMOS and I^2L technologies will continue to be directly competitive and most likely will coexist over the next several years for low- and medium-speed applications where a minimum number of chips and/or cost are of great importance (Ref. 7-2). Use of TTL and ECL technology as well as newer, fast, high-cost technologies will be restricted to high-performance systems, where the higher costs can be justified.

The use of LSI technology makes the CPUs smaller, by increased logic density on a chip; faster, because of the reduction in the propagation delay time between circuits; and inherently more reliable, by

reducing the number of solder joints and connections. The use of error-correcting codes in minicomputer memory systems will become standard and thus further increase system reliability.

Although bipolar main memories exist in some minicomputer systems and core main memories dominate today's installed base of minicomputers, the trend in new installations is clearly toward the less-expensive MOS technology. Continued demand for higher performance at a lower cost will bring MOS main memories into dominance. Main memory sizes will continue to increase. Virtual and cache memories will become standard on super-minicomputers and optional on most typical minicomputers.

Although any trends in minicomputer architecture are hard to discern, it is probable that bus-oriented structures will continue to be used through the early to mid-1980's. The single bus system will be especially important in very small systems which will have compatible microcomputers using the same bus system. In midi- and super-minicomputers, bus orientation will begin to yield its position to channel-oriented architectures now associated with large-scale computers by the early 1980's.

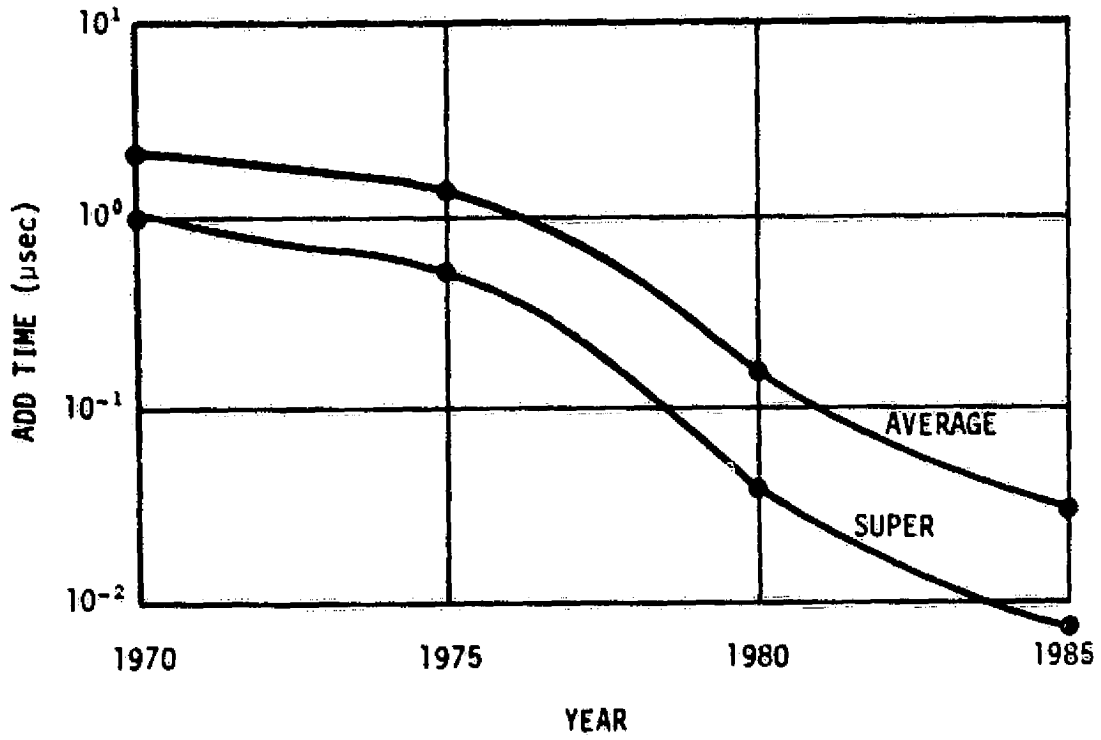


FIGURE 7.1.2.2-1. MINICOMPUTER ADD TIME TRENDS

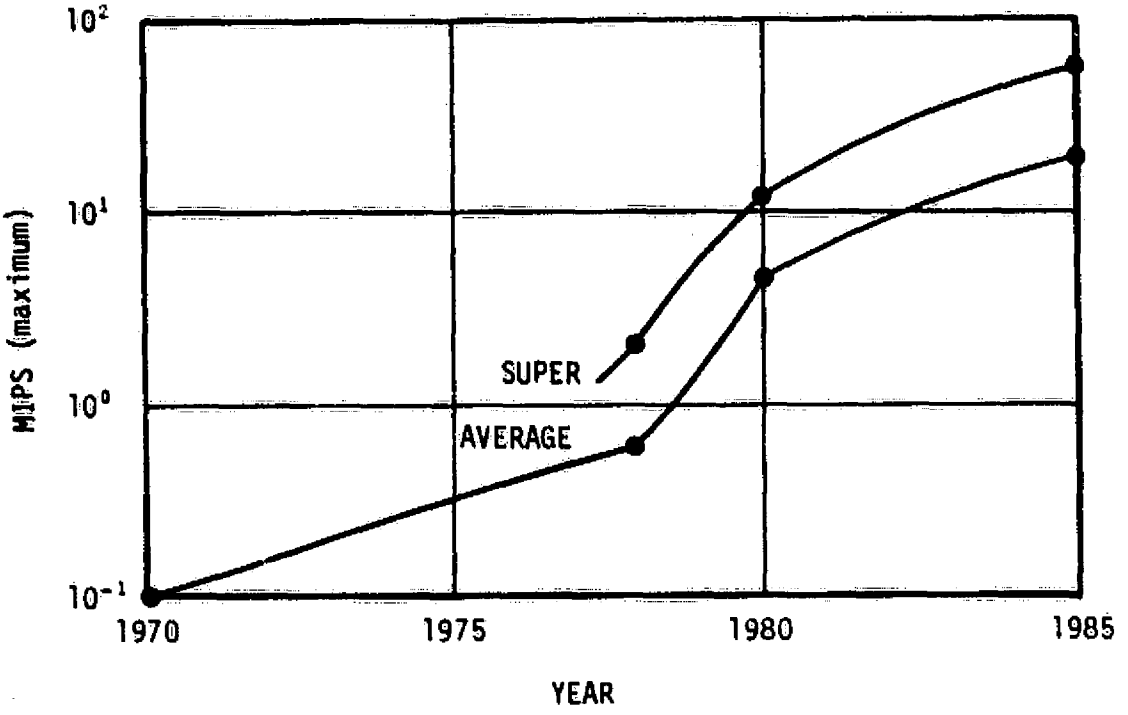


FIGURE 7.1.2.2-2. MINICOMPUTER INSTRUCTION EXECUTION RATE TRENDS

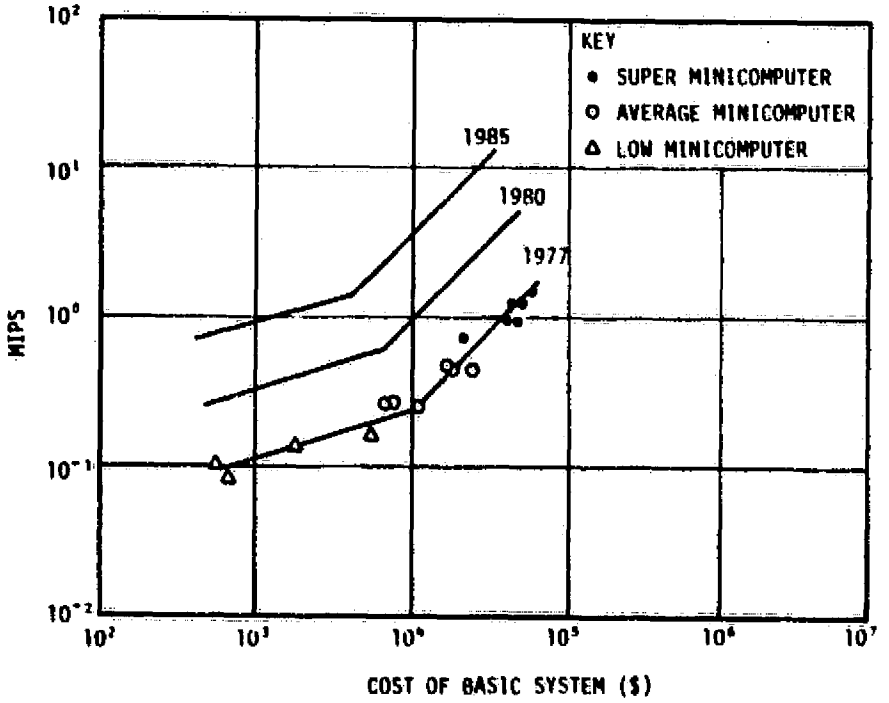


FIGURE 7.1.2.2-3. MINICOMPUTER COST TRENDS

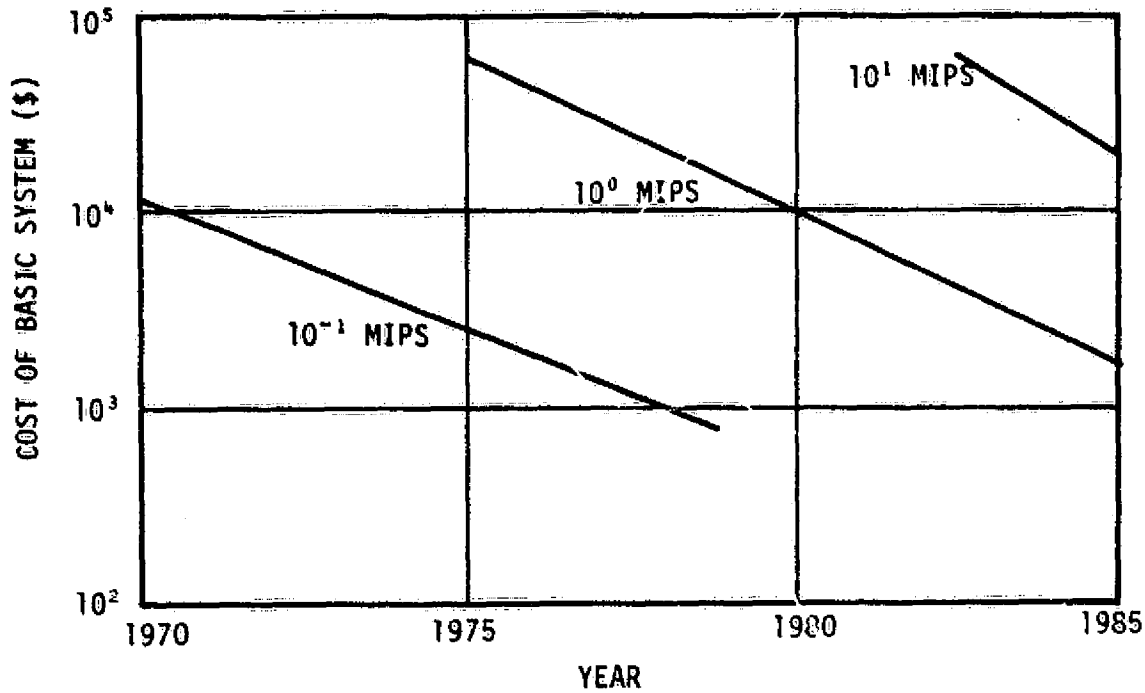


FIGURE 7.1.2.2-4. MINICOMPUTER COST TRENDS

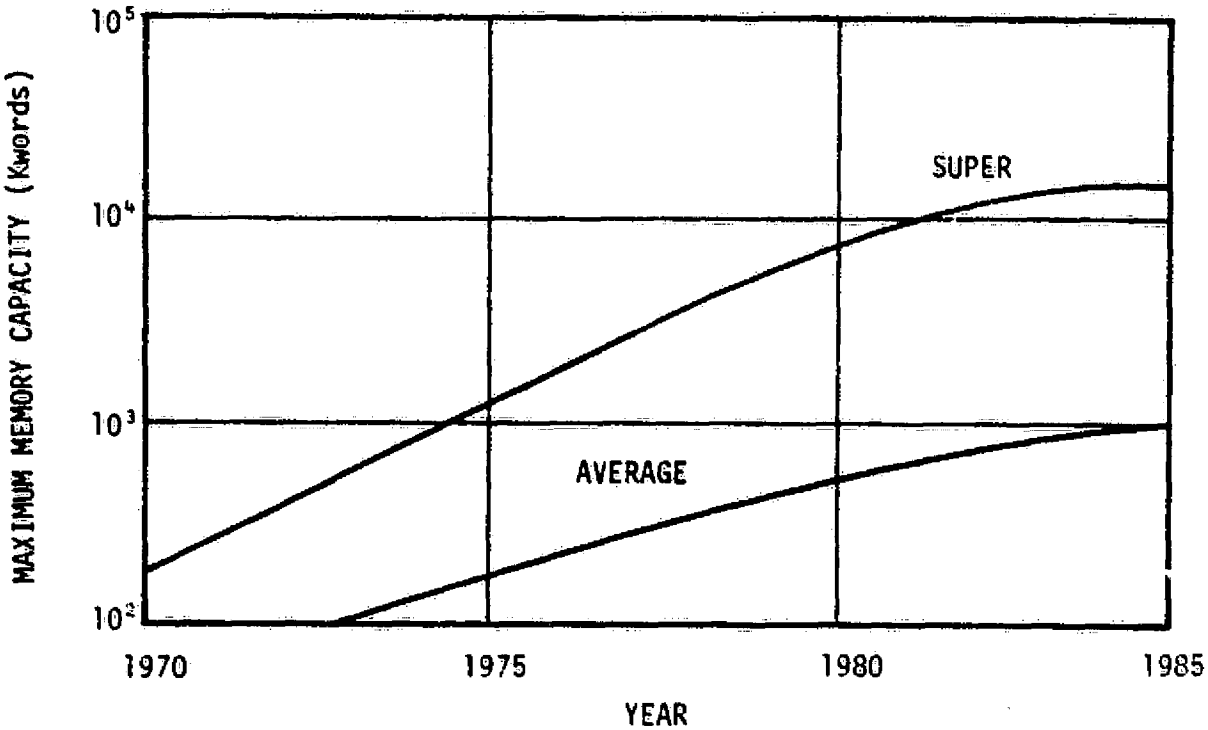


FIGURE 7.1.2.2-5. MINICOMPUTER MAXIMUM MEMORY CAPACITY TRENDS

7.1.2.3 Projected Developments in Minicomputers - Improvements in LSI logic circuits and semiconductor memories will continue to occur as a result of advancements in fabrication processes. Of major importance will be breakthroughs realized in the use of electron beams and X-ray lithography. These breakthroughs will be reflected in minicomputer performance improvements in many areas, including:

- CPU speed
- Main memory size and speed
- Cost
- Physical size
- Reliability
- Power requirements.

Tables 7.1.2.3-1 and 7.1.2.3-2 summarize the projected characteristics of minicomputers in 1980 and 1985, respectively. By the early 1980's, a complete minicomputer on a chip will appear, including a 16-bit CPU, 32 Kbits of memory, and simple I/O interfaces, with a manufacturing cost under $10 (Ref. 7-2). The cost of minicomputer systems utilizing these chips will be a function of peripherals, the volume purchased, and software capabilities provided with the system, since the hardware cost will be practically negligible.

The minicomputer of 1980 and 1985 will be able to support a fully capable data processing system, with ease of use in interactive applications being the primary design consideration (Ref. 7-3). This approach will force the typical minicomputer of 1985 to have a large main memory and a virtual operating system. By the early 1980's, hierarchical memory systems will be aided by the introduction of low-cost CCD-based or bubble memories for use as intermediate storage systems between main memory and disk storage. Physical volume and power requirements will drop to about 1/100 of today's systems in the 1980-1985 timeframe.

As the market addressed by minicomputers widens to include distributed processing and business data processing to a larger extent, greater flexibility of the basic minicomputer will be needed. To provide this flexibility, implementation of the instruction set in microprogrammed firmware rather than hardwired combinatorial logic will be more prevalent.

TABLE 7.1.2.3-1. PROJECTED DEVELOPMENTS FOR MINICOMPUTERS, 1980

CHARACTERISTICS	TYPICAL MINICOMPUTER	SUPER-MINICOMPUTER
CPU		
Word Length (bits)	16	16 or 32
Type of Logic Circuitry	NMOS, some I^2L	TTL or ECL
Cycle Time (μsec)	0.100 to 0.250	0.050 to 0.100
Add Time (μsec)	0.50	0.10
Number of Directly Addressable Words	NA	NA
Number of Instructions	NA	NA
Total Number of Registers	0 to 32	0 to 32
MIPS	0.8 to 1.2	2 to 4
Main Memory		
Type	Core or NMOS	NMOS
Cycle Time (μsec)	NA; 0.200 to 0.400	0.110 to 0.400
Minimum Capacity (words)	8K to 64K	32K to 128K
Maximum Capacity (words)	128K to 512K	256K to 8M
Word Length (bits)	16 plus parity for most; error correction for some systems	16 or 32 plus error correction for most systems
I/O Maximum Rate (words/sec)	NA	NA

TABLE 7.1.2.3-1 - Concluded

CHARACTERISTICS	TYPICAL MINICOMPUTER	SUPER-MINICOMPUTER
Software		
Assemblers	Assembler, macroassembler	Assembler, macroassembler
High-Level Languages	NA	NA
Monitors/Operating Systems	NA	NA
Physical Characteristics		
Volume (in^3)	3,000	30,000
Voltage (Vac)	115 at 60 Hz or 220 at 50 Hz	115 at 60 Hz or 220 at 50 Hz
Operating Temperature Range (°C)	0 to 50	0 to 50
Humidity Tolerance (%)	0 to 95	0 to 95
Cost of Basic System (CPU, power supply, front panel, and minimum memory in a chassis)	$7,000 to $18,000	$20,000 to $75,000
MTBF (hr)	NA	NA
Comments	Virtual memory and cache memory on some systems	Virtual memory and cache memory on many systems

TABLE 7.1.2.3-2. PROJECTED DEVELOPMENTS FOR MINICOMPUTERS, 1985

CHARACTERISTICS	TYPICAL MINICOMPUTER	SUPER-MINICOMPUTER
CPU		
Word Length (bits)	16 or 32	32
Type of Logic Circuitry	NA	NA
Cycle Time (μsec)	0.010 to 0.015	0.004 to 0.008
Add Time (μsec)	0.03	0.008
Number of Directly Addressable Words	NA	NA
Number of Instructions	NA	NA
Total Number of Registers	NA	NA
MIPS	10 to 20	30 to 60
Main Memory		
Type	NMOS	NMOS or Bipolar
Cycle Time (μsec)	0.100 to 0.250	0.080 to 0.250; 0.055 to 0.085
Minimum Capacity (words)	16K to 128K	64K to 256K
Maximum Capacity (words)	256K to 1M	1M to 16M
Word Length (bits)	16 or 32 plus error correction for many systems	32 plus error correction
I/O Maximum Rate (words/sec)	NA	NA

TABLE 7.1.2.3-2 - Concluded

CHARACTERISTICS	TYPICAL MINICOMPUTER	SUPER-MINICOMPUTER
Software		
Assemblers	Assembler, macroassembler	Assembler, macroassembler
High-Level Languages	NA	NA
Monitors/Operating Systems	NA	NA
Physical Characteristics		
Volume (in^3)	50	500
Voltage (Vac)	115 at 60 Hz or 220 at 50 Hz	115 at 60 Hz or 220 at 50 Hz
Operating Temperature Range (°C)	0 to 50	0 to 50
Humidity Tolerance (%)	0 to 95	0 to 95
Cost of Basic System (CPU, power supply, front panel, and minimum memory in a chassis)	$10,000 to $15,000	$20,000 to $40,000
MTBF (hr)	10,000	10,000
Comments	Virtual memory and cache memory on many systems	Virtual memory and cache memory on most systems

Channel-oriented architectures used in large-scale computers of today will be found in some super-minicomputers of 1980 and will be commonplace by 1985. Bus-oriented architectures will continue to remain popular in the lower and middle ranges of minicomputers even in 1985. But even here, channel-oriented architectures will begin to make inroads.

7.1.3 Large-Scale Computers

7.1.3.1 State of the Art in Large-Scale Computers - A summary of the basic characteristics of the most powerful current large-scale computers is given in Table 7.1.3.1-1. The cost of a basic system (CPU, I/O channels, system console, power distribution unit, and minimum amount of memory) varies between $3 million and $5 million, with cost generally increasing with speed and computing power. The high performance levels of recent vintage large-scale systems, especially their speeds, have been achieved through the use of integrated logic circuitry, semiconductor main memories, buffer or cache memories, instruction look-ahead, concurrent CPU and I/O processing, and multiprocessing. Advancements in these areas have reduced significantly the price/performance ratio for large-scale systems, as reflected in the low cost (0.5¢) to perform 100,000 computations.

Although discrete components still appear in a very few CPUs, most now utilize integrated circuits with ECL technology. The result has been faster and inherently more reliable CPUs. Some models incorporate microprogrammed control logic for added flexibility.

Core main memories have not been totally eliminated, but the use of semiconductor main memories is now predominant. As a result, the physical volume occupied by the main memory in a system for a comparable number of storage locations has decreased greatly. Although both bipolar and MOS technologies are in use today, more systems use MOS memory because it is by far more cost-effective. The major advantage of bipolar over MOS is greater speed, but this is obtained at the expense of higher cost, much greater power consumption, higher heat dissipation, and lower packing densities. Memory interleaving techniques are used to increase effective speed so that bipolar's speed is unnecessary in most cases. The current large-scale computers have maximum main memory capacities up to 16 Mbytes. Virtual storage capabilities, available in about half of these computers, further enlarge the effective memory capacity, from the user's viewpoint, but the overhead required may introduce undesirable delays.

TABLE 7.1.3.1-1. STATE OF THE ART IN LARGE-SCALE COMPUTERS

CHARACTERISTICS	
CPU	
Word Length (bits)	32 or 48 or 60
Type of Logic Circuitry	ECL
Cycle Time (nsec)	25 to 80
Add Time:	
Fixed Point (μsec)	0.055 to 0.13
Floating Point (μsec)	0.110 to 0.7
Number of Instructions*	100 to 200
Total Number of Registers	8 to 64
MIPS	10 to 25
Main Memory	
Type	MOS or Core
Cycle Time (nsec)	300 to 500; 750 to 900
Minimum Capacity (bytes)	1M to 2M
Maximum Capacity (bytes)	4M to 16M
Word Length (bits)	Same as CPU plus parity or error correction on most systems
Cache Memory	
Type	TTL or ECL
Cycle Time (nsec)	32 to 80
Capacity (bytes)	16K to 32K
Virtual Memory	Some systems
I/O Maximum Rate (bytes/sec)	4M to 18M

*Instruction sets are very comprehensive, including capabilities for fixed and floating point decimal and binary arithmetic and string manipulations, among others.

TABLE 7.1.3.1-1 - Concluded

CHARACTERISTICS	
Software	
Assemblers	Macroassembler
High-Level Languages	FORTRAN, COBOL, BASIC, ALGOL, PL/1, etc.
Monitors/Operating Systems	Complex batch, time-sharing, etc., operating systems
Cost of Basic System (CPU, I/O channels, system console, power distribution unit, and minimum memory)	$3M to $5M
Cost per 100,000 Computations (¢)	0.5
MTBF (hr)	150 to 250

Buffer or cache memories are included in many large-scale computers. This relatively small but very fast memory is provided to optimize the processing capability of the CPU. The most frequently referenced items are kept in this fast memory, which causes a decrease in the average memory reference time. For the same purpose, instruction look-ahead is used, either in conjunction with or instead of buffer memory. This technique increases CPU utilization by minimizing the time required for memory references needed to access sequences of instructions.

All of the state-of-the-art large-scale computers have separate I/O processors that operate independently of the CPU. This allows concurrent data transmission and CPU execution, yielding an increase in CPU utilization and system throughput. Multiprocessing, i.e., the use of more than one CPU in the system, achieves even greater increases in throughput through concurrent execution by the CPUs. Several of the current large-scale computers are designed to operate in a multiprocessing mode. The instruction overlap or streaming structure (sometimes erroneously referred to as pipelining) found in some CPUs lies at an intermediate level, allowing several instructions to be in some phase of execution concurrently within a single CPU.

More emphasis has been placed on reliability in the current large-scale systems. The use of integrated circuits makes the logic circuitry inherently more reliable because of reduced component count and fewer solder joints and connections. Algebraic codes incorporated in the data stream increase the reliability of data transmission. Code bits within each memory word allow correction of single-bit errors and detection of all multiple errors. Extensive parity checking is performed on instructions and I/O data. Most systems have an automatic instruction re-try to correct transient errors. Fail-soft capabilities are being emphasized, enabling a system to remain in operation at a reduced level of performance when failures occur. As a result, system reliability with a mean time between failure (MTBF) of 150 to 250 hr is now typical.

The instruction sets of the large-scale computers are generally very comprehensive. Most repertoires include instructions for data movement, shifting, logic and control operations, binary arithmetic, variable-length fixed and floating point decimal arithmetic, and bit and byte string manipulation. The most powerful computers permit floating point and integer operands of single and double precision to be combined freely, and also include vector mode operations.

Almost all current large-scale computer systems use a channel-oriented architecture as shown in Figure 7.1.3.1-1. Many variations on this basic architecture exist, including number and use of buffers, multiple processor units (CPUs), multiple memory units, several levels of memory, a large variation in the number and types of I/O channels, and some degree of parallelism within modules. In general, the power of the computer is higher in designs that incorporate more of these features. As an alternative architecture, at least one computer manufacturer has used a bus-oriented structure in its midscale computers. For a discussion of bus-oriented architectures, see Subsection 7.1.2.1.

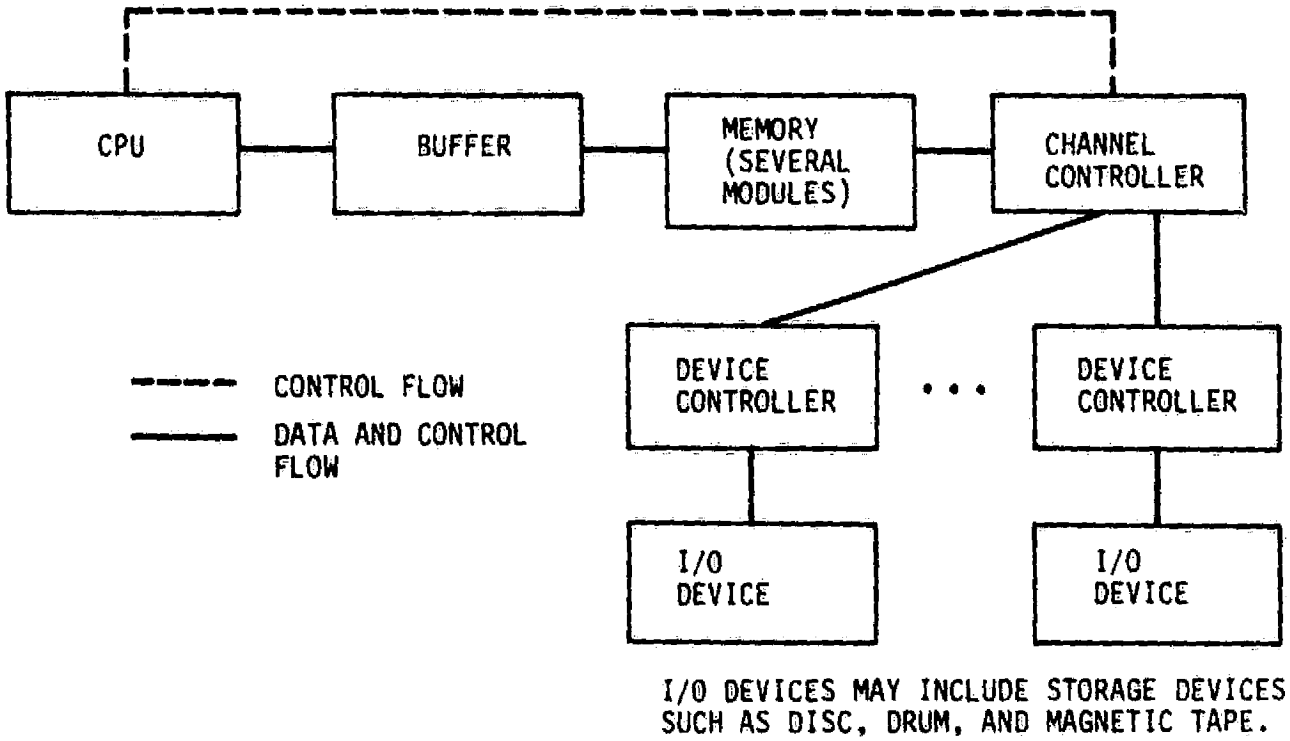


FIGURE 7.1.3.1-1. LARGE-SCALE COMPUTER ARCHITECTURE

7.1.3.2 Trends in Large-Scale Computers - During the past decade, the capabilities of large-scale computers have increased rapidly while their costs per MIPS have decreased. Technological and manufacturing innovations will continue these trends over the next several years, as illustrated graphically in Figures 7.1.3.2-1 and 7.1.3.2-2. Although the graph depicting MIPS as a function of year (Figure 7.1.3.2-1) was originally produced in 1972, its predictions for today's large-scale computers are still valid and those for large-scale computers of the 1980's are consistent with more recent predictions. The historical 12 to 15% annual price/performance improvement will continue as the result of achievements in logic circuits, main memories, and mass storage (Ref. 7-4). Reflecting this trend, the cost to perform 100,000 computations will continue to decline, as illustrated graphically in Figure 7.1.3.2-3.

There is a definite trend toward the use of large-scale integrated (LSI) logic circuitry in large-scale computers. In the future, only LSI-compatible logic types will be used. Denser logic chips will continue to be achieved through research and developments in scanning techniques, projection techniques, electron beam technology, and X-ray technology. Improved system speed and reduced system power dissipation will be achieved, both through the use of these denser chips and the use of newer logic circuit technology such as Josephson junctions, GaAs, or short-channel MOS (VMOS, DMOS, etc.). The use of microprogram control logic will become more pronounced and more sophisticated.

During the next few years, emphasis on system reliability will increase. Fault-tolerant design with redundancy at the module level will be aided by the increasing availability of larger and denser LSI devices. There will be continued widespread use of on-line monitoring and diagnostic capabilities with automatic re-try and path switching. The use of error correcting codes in memory systems will grow.

Although core main memories still exist in some of today's installed large-scale computers, the trend is clearly toward semiconductor main

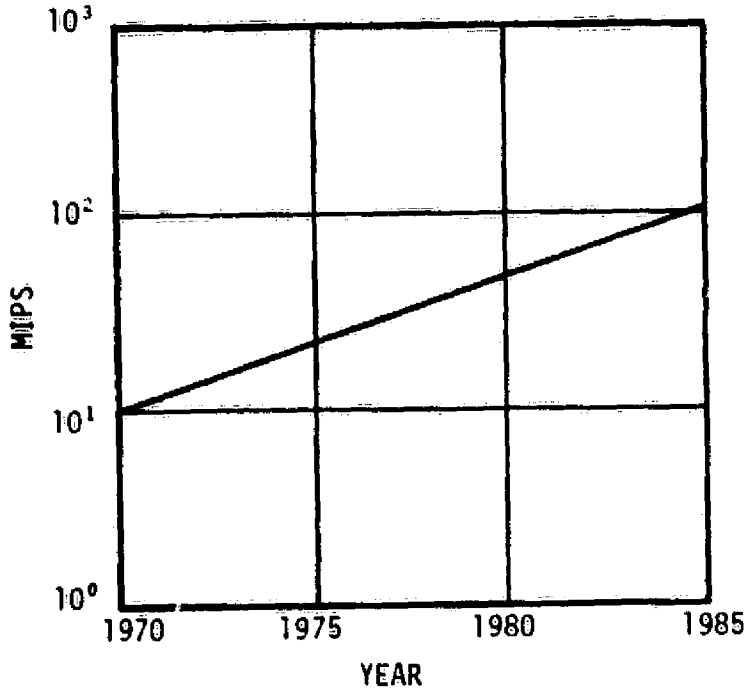


FIGURE 7.1.3.2-1. LARGE-SCALE COMPUTER INSTRUCTION EXECUTION RATE TRENDS (Ref. 7-5)

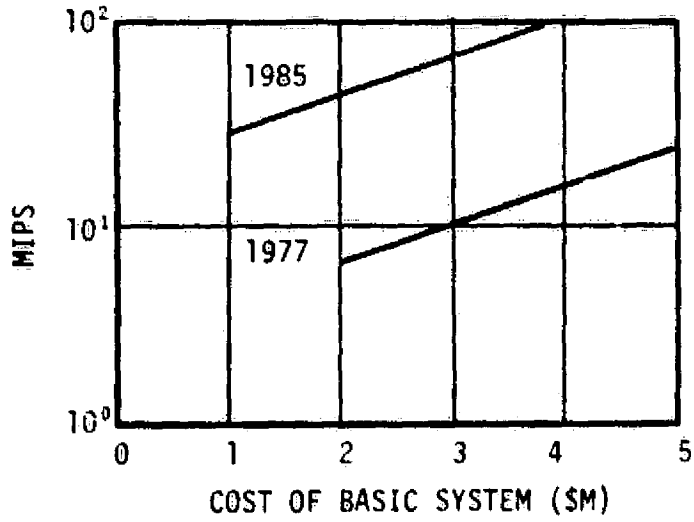


FIGURE 7.1.3.2-2. LARGE-SCALE COMPUTER COST TRENDS

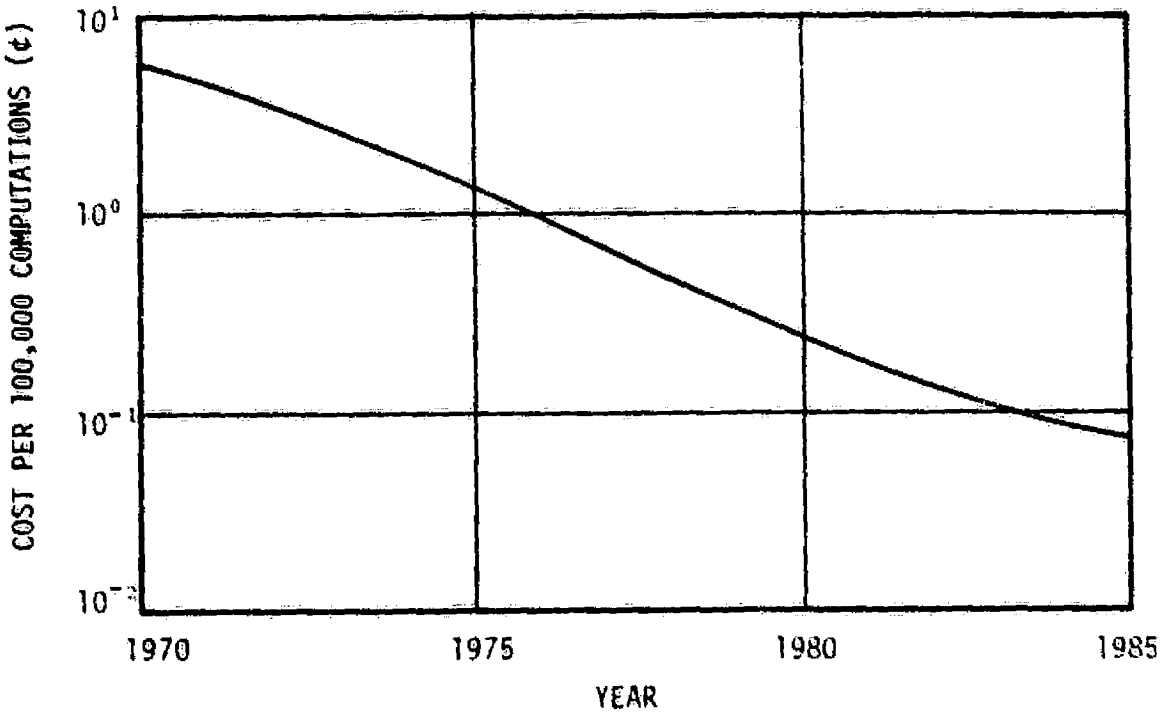


FIGURE 7.1.3.2-3. LARGE-SCALE COMPUTER COST PER 100,000 COMPUTATIONS TRENDS (Ref. 7-6)

memories. Main memory capacities will continue to increase while their physical dimensions will continue to decrease. These trends are illustrated graphically in Figures 7.1.3.2-4 and 7.1.3.2-5.

The desire for ever greater speed will force the continued growth of multilevel hierarchical storage systems. In particular, buffer memories will be included in strategic locations throughout large-scale computers and will have larger capacities and smaller cycle times.

The trends in large-scale computing include a move in some applications from single-machine computers toward distributed processing, where dedicated, interconnected processors work on each segment of a task (Ref. 7-7). Much research is being done in this area. Such systems are projected by many experts for applications having specific problems such as speed or data organization. As support for their positions, these

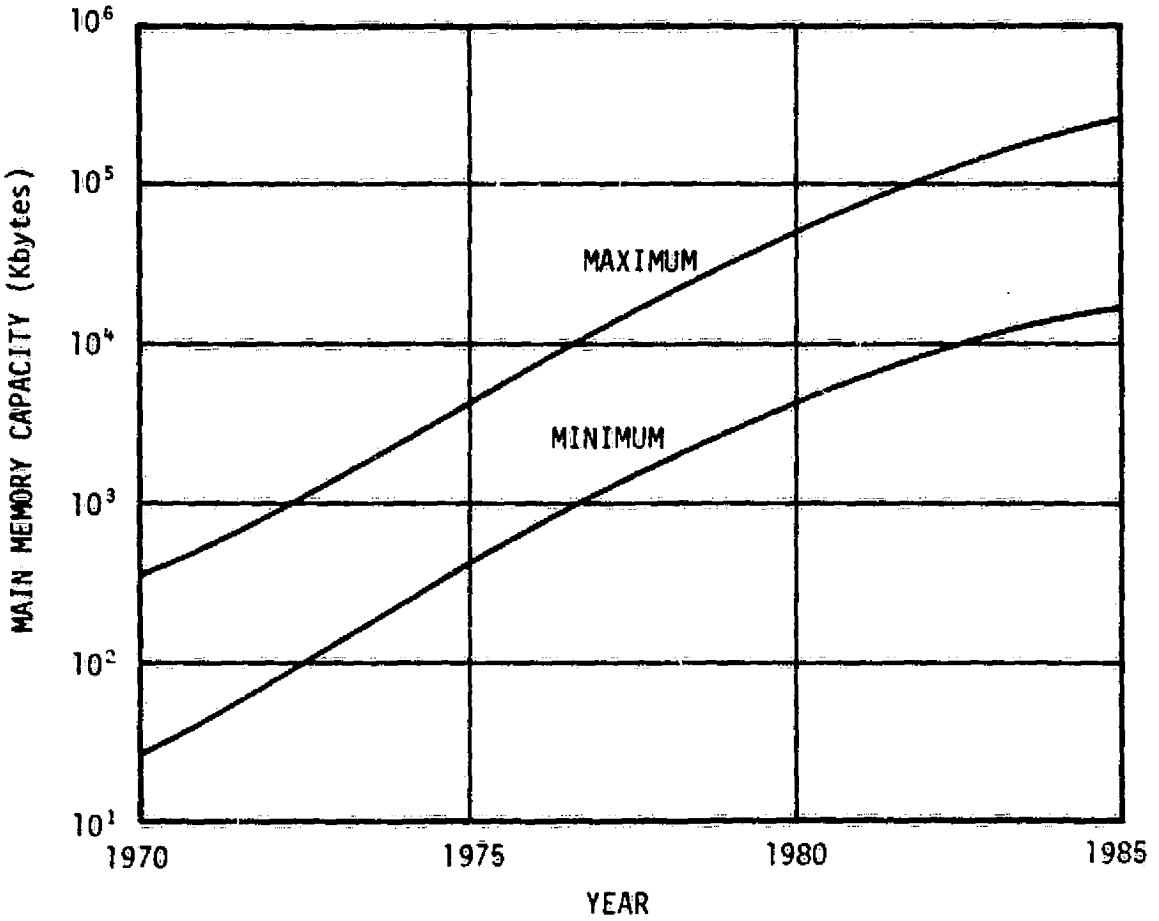


FIGURE 7.1.3.2-4. LARGE-SCALE COMPUTER MAIN MEMORY CAPACITY TRENDS

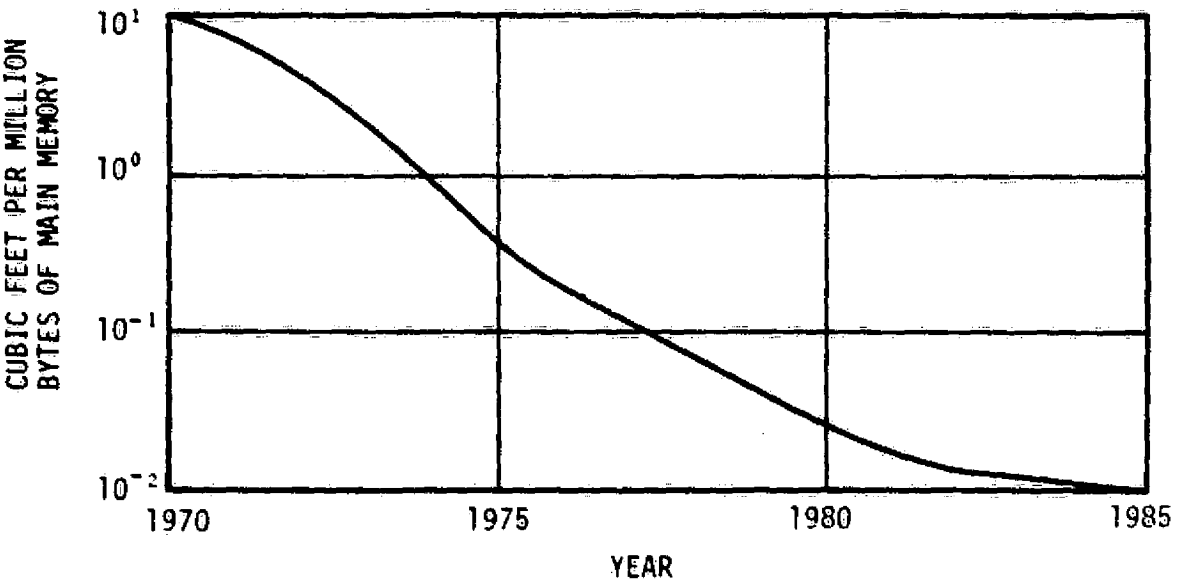


FIGURE 7.1.3.2-5. LARGE-SCALE COMPUTER MAIN MEMORY VOLUME TREND (Ref. 7-6)

experts cite the increasing complexity of the large computer systems, particularly the software required to utilize their vast array of features, and the decreasing cost and increasing capabilities of minicomputers and microcomputers. Although distributed processing has progressed only to the first stage, where a small amount of processing has been taken away from the CPU, a greater degree of off-loading of the CPU can be expected (Ref. 7-8).

Some trends in architecture for large-scale computers are toward increasing use of multiple CPUs, more and larger buffers, and larger memory. Additionally, more parallelism within modules (borrowed from current super-scale computer architecture) is being incorporated in large-scale computers. A somewhat counter trend toward bus-oriented architectures is weakly evident in midscale computers but is not likely to be very important through the mid 1980's.

7.1.3.3 Projected Developments in Large-Scale Computers - Improvements in LSI logic circuits and semiconductor memories will continue to advance (see Subsection 7.2.1). These improvements will be reflected in improvements in large-scale computers in many areas, including:

- CPU speed
- Main memory size and speed
- Buffer memory size and speed
- Cost
- Reliability.

Tables 7.1.3.3-1 and 7.1.3.3-2 summarize the projected characteristics of large-scale computers in 1980 and 1985, respectively. Bipolar logic circuits will continue to be dominant in the CPUs of large-scale computers because speed is more critical than packing density for these machines. Microcoded instruction sets will be used extensively in the 1980's for flexibility and protection against plug-compatible devices.

Hierarchical storage systems with larger, faster, and more intelligent buffer memories will be dominant in the large-scale systems by 1985 to aid in achieving desired system performance. The semiconductor main memories of the hierarchy will have capacities up to 256 Mbytes in 1985, while virtual address spaces of 10^{12} bytes are likely by that time. By 1980, low-cost CCD-based or bubble memories which have already begun appearing in the hierarchy of some superscale computers (Ref. 7-9) will be available as secondary storage on large-scale machines.

Error detection and error correction in memory systems and fault-tolerant circuitry will become standard, thus increasing the reliability of large-scale systems. Multiple-bit error correction in memory systems will become available. Multiprocessors will be included in some systems, primarily for increased system reliability.

Distributed processing systems will increase in number through the 1980's, with increasing amounts of the processing off-loaded from the CPU. By 1980, the distributed processing network era will have arrived (Ref. 7-6). However, this progress will be in conjunction with,

TABLE 7.1.3.3-1. PROJECTED TECHNOLOGY FOR LARGE-SCALE COMPUTERS, 1980

CHARACTERISTICS	
CPU	
Word Length (bits)	32 or 48 or 60 or 64
Type of Logic Circuitry	ECL, some CML
Cycle Time (nsec)	18 to 58
Add Time:	
Fixed Point (μsec)	0.04 to 0.09
Floating Point (μsec)	0.08 to 0.51
Number of Instructions	NA
Total Number of Registers	NA
MIPS	40 to 45
Main Memory	
Type	MOS
Cycle Time (nsec)	110 to 300
Minimum Capacity (bytes)	4M
Maximum Capacity (bytes)	64M
Word Length (bits)	Same as CPU plus error correction on most systems
Cache Memory	
Type	TTL or ECL
Cycle Time (nsec)	26 to 64
Capacity (bytes)	32K to 64K
Virtual Memory	Some systems
I/O Maximum I/O Rate (bytes/sec)	NA
Software	
Assemblers	Macroassembler
High-Level Languages	FORTRAN, COBOL, BASIC, ALGOL, PL/1, etc.
Monitors/Operating Systems	NA

TABLE 7.1.3.3-1 - Concluded

CHARACTERISTICS	
Cost of Basic System (CPU, I/O channels, system console, power distribution unit, and minimum memory)	$2.6M to $4.5M
Cost per 100,000 Computations (¢)	0.22
MTBF (hr)	350 to 650

TABLE 7.1.3.3-2. PROJECTED TECHNOLOGY FOR LARGE-SCALE COMPUTERS 1985 (Refs. 7-9 and 7-10)

CHARACTERISTICS	
CPU	
Word Length (bits)	32 or 48 or 60 or 64
Type of Logic Circuitry	TBD
Cycle Time (nsec)	5 to 32
Add Time:	
Fixed Point (μsec)	0.005 to 0.032
Floating Point (μsec)	0.04 to 0.28
Number of Instructions	NA
Total Number of Registers	NA
MIPS	90 to 100
Main Memory	
Type	MOS or Bipolar
Cycle Time (nsec)	50 to 100
Minimum Capacity (bytes)	17.4M
Maximum Capacity (bytes)	256M
Word Length (bits)	Same as CPU plus error correction
Cache Memory	
Type	TBD
Cycle Time (nsec)	20
Capacity (bytes)	128K to 256K
Virtual Memory	Many systems
I/O Maximum Rate (bytes/sec)	NA
Software	
Assemblers	Macroassembler
High-Level Languages	FORTRAN, COBOL, BASIC ALGOL, PL/1, etc.
Monitors/Operating Systems	NA

TABLE 7.1.3.3-2 - Concluded

CHARACTERISTICS	
Cost of Basic System (CPU, I/O channels, system console, power distribution unit, and minimum memory)	$1.8M to $3.75M
Cost per 100,000 Computations (¢)	0.08
MTBF (hr)	1.500 to 2.500

rather than instead of, the evolution of large-scale computers. In addition, many large-scale systems of the 1980's will be geared to large data-base management operations.

Large-scale computer architectures of the 1980's will be very similar to today's architectures, with additional emphasis on multiple CPUs, additional buffers, and increased parallelism. By 1985, these additions will be more pronounced and features to support distributed processing will be incorporated.

7.1.4 Super-Scale Computers

7.1.4.1 State of the Art in Super-Scale Computers - Computers that execute 50 MIPS or more are generally classified as super-scale computers at the present time. The super-scale computer state of the art encompasses three basic designs: vector or pipeline processors, parallel array processors, and associative array processors. All three types are designed primarily to solve problems that throughput large amounts of data and in which computations can be structured in a highly parallel manner, such as weather forecasting, sonar processing, air traffic control, and image processing. Nonparallel (sequential) problems may also be handled by these machines, but with a great reduction in execution rates. For problems of a highly sequential nature that still demand very high throughput, large-scale multiprocessor computers are a viable alternative to super-scale computers.

The architecture of vector super-scale computers is very similar to that of large-scale computers, with a few additions. Vector super-scale computers are characterized by a single data stream that passes through a pipeline of sequentially connected processing units. Each processing unit operates on the elements of the data stream sequentially. Once results are first obtained from the pipe, they will be produced at the rate at which it takes a single processor of the pipe to perform one operation. Thus the high instruction execution rates claimed for these computers are based on the assumptions that there is a long string of data, i.e., a vector, which needs to be processed by the same sequence of operations, and that these data items can be fetched from memory at speeds corresponding to the processor speed. A few of these pipelines may be paralleled. Additionally, the I/O facilities of these computers are often structured more for communication with smaller computers (used as front-end processors) than for communication with user-oriented devices.

A summary of the basic characteristics of state-of-the-art vector super-scale computers is presented in Table 7.1.4.1-1. Although scalar

processing capabilities are included in these systems, emphasis is placed on vector processing. It is through the utilization of the vector capabilities that extremely high instruction execution rates are achieved. The rates included in Table 7.1.4.1-1 are the maximum rates possible. In actual applications, the rates achieved are about one-third to one-half of these maximums. To enable the system to operate near the high vector rates, the vector super-scale computers contain the following features:

- Bipolar integrated logic circuits
- Bipolar main memories
- Instruction stack with look-ahead hardware
- Independent I/O processors
- I/O channel buffering capabilities.

The newest systems use ECL logic circuits and memories because of the speed advantage of this technology over other semiconductor technologies. The high main memory speeds and large main memory capacities enable these systems to handle the large amounts of data inherent in the problems they address at rates commensurate with the vector rates. All of these systems include separate I/O processors operating independently of the CPU. This feature allows concurrent data transmission and CPU execution, yielding optimized CPU utilization. Several of these systems utilize a separate large-scale mainframe acting as a host to the vector super-scale computer. This configuration off-loads most of the control tasks and some of the sequential tasks from the vector computer, allowing the vector computer to execute primarily in its fast vector mode.

Architecturally, a parallel array processor is a single-instruction-stream, multiple-data-stream computer. It contains many independent identical processing elements, each operating on a separate data stream but all under control of the single instruction stream. An associative array processor is just a special class of parallel array processor in which the processing elements are simple, serial-by-bit units connected to an associative memory, one unit per row of memory. Parallelism is achieved by performing the same operation on all words of the associative memory concurrently. The high instruction execution rates claimed for array processor super-scale

TABLE 7.1.4.1-1. STATE OF THE ART IN VECTOR SUPER-SCALE COMPUTERS

CHARACTERISTICS	
CPU	
Word Length (bits)	32 or 64
Type of Logic Circuitry	TTL or ECL
Cycle Time (nsec)	12.5 to 40
Minimum Computation Times:	
64-bit Add (nsec)	12.5 to 20
64-bit Multiply (nsec)	12.5 to 48
64-bit Divide (nsec)	40 to 80
Maximum Vector Result Rate (nsec)	12.5 to 40
Number of Instructions:	
Scalar	64 to 177
Vector	64 to 70
Total Number of Registers	Tens to hundreds
MIPS	50 to 80
Memory	
Type	TTL or ECL
Cycle Time (nsec)	50 to 160
Maximum Capacity (words)	1M
Bandwidth (Mwords/sec)	80 to 400
Error Correction	Single error correction on some systems
Virtual Memory	Only on a few systems
I/O Maximum Rate (bytes/sec)	To 400M
Software	
Assemblers	Macroassembler
High-Level Languages	FORTRAN
Monitors/Operating Systems	Batch operating system

TABLE 7.1.4.1-1 - Concluded

CHARACTERISTICS	
Physical Characteristics	
Volume (ft^3)	170 to 450
Mainframe Weight (tons)	5 to 20
Cost ($M)	7.5 to 12.5
MTBF (hr)	2.5 to 5 (no memory error correction)
	40 to 60 (memory error correction)
Maximum Number of Pipelines	1, 2, or 4

computers are based on the concurrent operation of all processing elements, which in turn is dependent on the availability of very large amounts of data requiring identical processing. To an even greater extent than in vector super-scale computers, the performance of these computers is seriously degraded by the need for scalar processing.

A summary of the basic characteristics of state-of-the-art parallel and associative array processors is presented in Table 7.1.4.1-2. The instruction execution rates included in the table are the maximum rates possible. In actual applications, the rates achieved are about one-third to one-half of these maximums. The newest systems use ECL logic circuits and processing element memories because of the speed advantage of this technology over other semiconductor technologies. However, the individual processing elements are not extremely powerful and do not have large amounts of memory. It is the concurrent operation of these processing elements that yields a high-performance system. A large-scale mainframe is included in the system to execute control tasks and many of the sequential tasks, allowing the processing elements to operate in parallel as much as possible.

Super-scale computers are relatively unreliable because they include massive amounts of hardware concentrated in one place. A number of features have been included in both types of systems to aid reliability, including:

- Error-correcting memories
- Automatic instruction re-try
- Integrated logic circuits.

Array processors have the added advantage that an element failure can usually be tolerated and frequently even ignored in the kinds of problems addressed by them.

Software support for the super-scale computers is fairly limited. It generally includes an operating system, an assembly language, FORTRAN or a FORTRAN-like higher-level language, a library of mathematical routines, and a general utilities package. The operating system overhead is kept as low as possible because computers designed for heavy computational workloads

TABLE 7.1.4.1-2. STATE OF THE ART IN ARRAY SUPER-SCALE COMPUTERS

CHARACTERISTICS	PARALLEL	ASSOCIATIVE
Processing Elements		
Number per System	64 to 288	256 per module; up to 32 modules
Word Length (bits)	32 or 64	32
Type of Logic Circuitry	ECL	NA
Cycle Time (nsec)	75 to 80	NA
Processing Element Data Memory		
Type	NMOS or ECL	NMOS
Cycle Time (nsec)	100 to 240	200
Maximum Capacity	2K words	256 bits/processing element; 64 Kbits/ module
Error Correction	Yes	Yes
MIPS	1 to 4	NA
System MIPS	100 to 500	53.3/module**
Program and Data Memory*		
Type	NMOS	NMOS
Cycle Time (nsec)	200	200
Maximum Capacity (words)	32K	4K***
Error Correction	Yes	Yes
I/O Maximum Rate	Relative to maximum rates of external devices	Relative to maximum rates of external devices
Software	Minimal	Minimal
System Cost ($M)	10	0.7 to 3.5
System Up-Time (hr/5-day wk)	60	NA

*On some machines
**Word search operations
***Also 32K words of bulk core memory

function most efficiently under these circumstances. Much work is needed in the area of software support, especially support to allow better utilization of the vector or parallel capabilities of the super-scale computers.

7.1.4.2 Trends in Super-Scale Computers - Historically, the performance levels of super-scale computers have increased rapidly while their costs have increased at a slower rate, resulting in a continuous price/performance improvement. Technological and manufacturing innovations in the areas of logic circuits, main memories, and mass storage will continue these trends into the 1980's, as illustrated graphically in Figures 7.1.4.2-1 through 7.1.4.2-4. These figures, as well as the other figures contained in this subsection, are intended to illustrate trends only and should not be used to extract numbers for any particular year. Because of the limited market for super-scale computers and their high development cost and level of complexity, the lead time for the production of a new system is long. As a result, improvements occur in jumps spaced several years apart rather than in small increments every year or so. For a given year, the curves indicate what the capabilities would probably be if a new system were to be introduced that year. It is probable that only one new generation of array super-scale computers will appear during the 1980-1985 timeframe.

Emphasis will be placed on larger memories, faster execution, and improved software in the super-scale computers in the 1980's (without changing the basic architecture of the machine). The same problems will continue to be addressed and the quality of their solution depends on the number of data points considered and the number of computations performed. Figures 7.1.4.2-5 and 7.1.4.2-6 graphically illustrate the expected growth in memory capacities. ECL technology will soon be replaced by advanced technologies that provide both greater speed and much lower power and are thus compatible with LSI circuit technology. Some potential future logic candidates include Schottky GaAs MESFETS, Josephson junctions, short-channel MOS devices such as VMOS and DMOS, and elevated electrode logic. Most of these technologies are discussed in Section 1. Denser logic chips and memory chips will be achieved. These achievements are basic technology improvements and are reflected throughout the computer industry -- not constrained to a particular group of computers. Improved

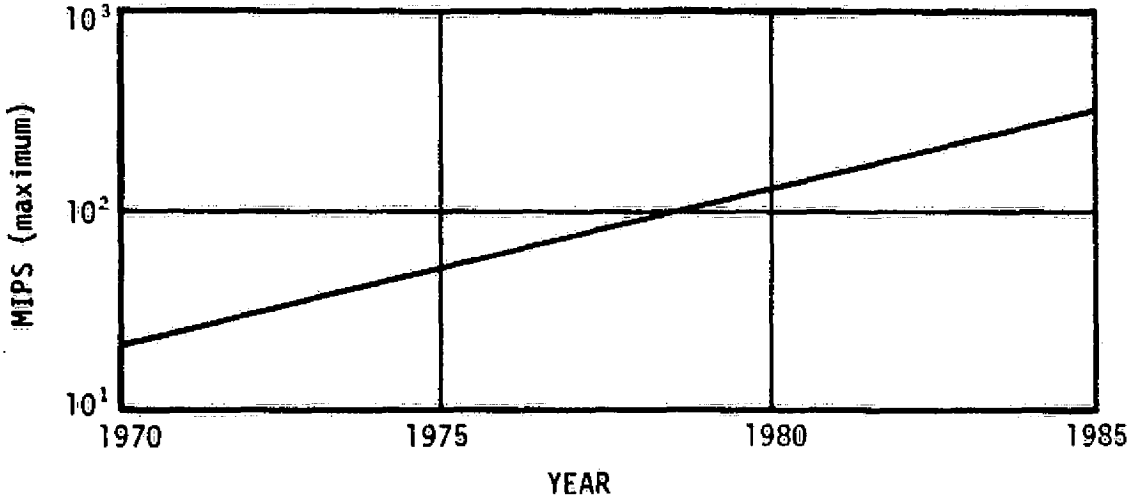


FIGURE 7.1.4.2-1. VECTOR SUPER-SCALE COMPUTER INSTRUCTION EXECUTION RATE TREND

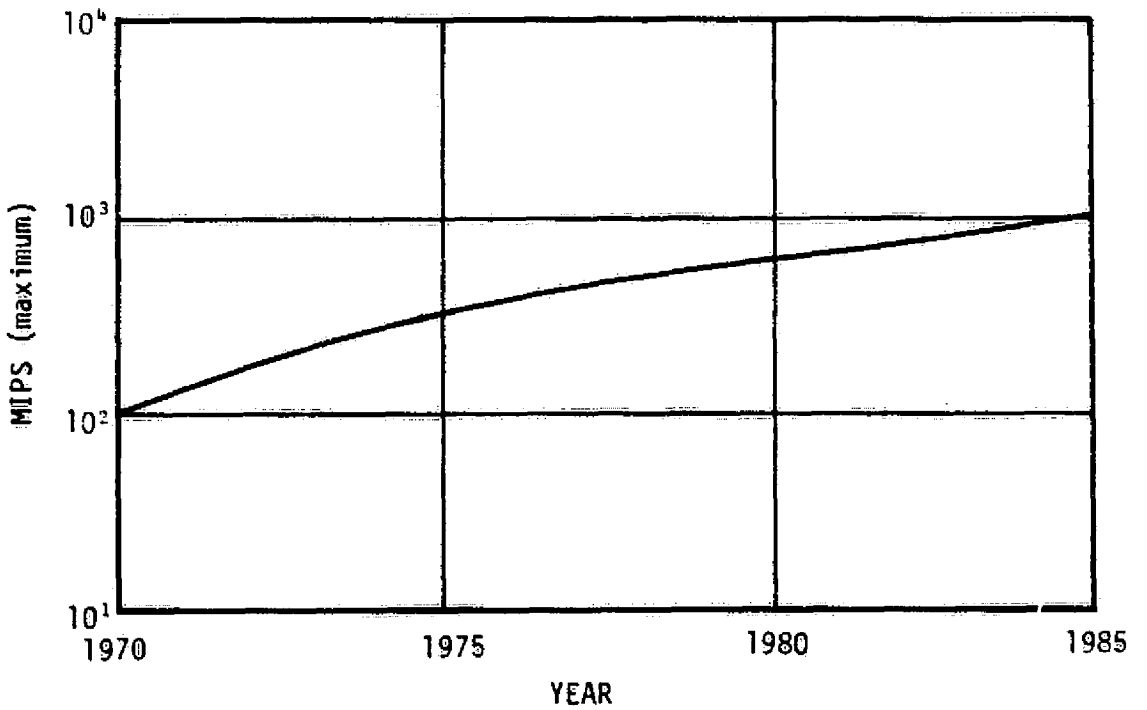


FIGURE 7.1.4.2-2. ARRAY SUPER-SCALE COMPUTER INSTRUCTION EXECUTION RATE TREND

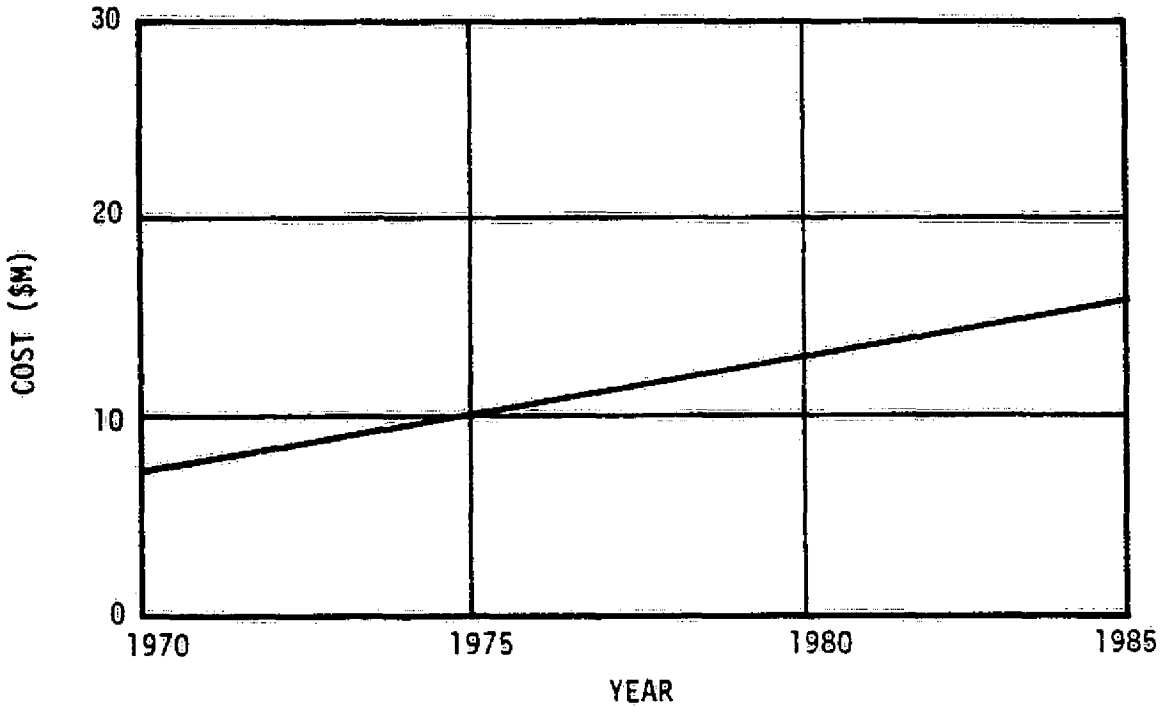


FIGURE 7.1.4.2-3. VECTOR SUPER-SCALE COMPUTER COST TREND

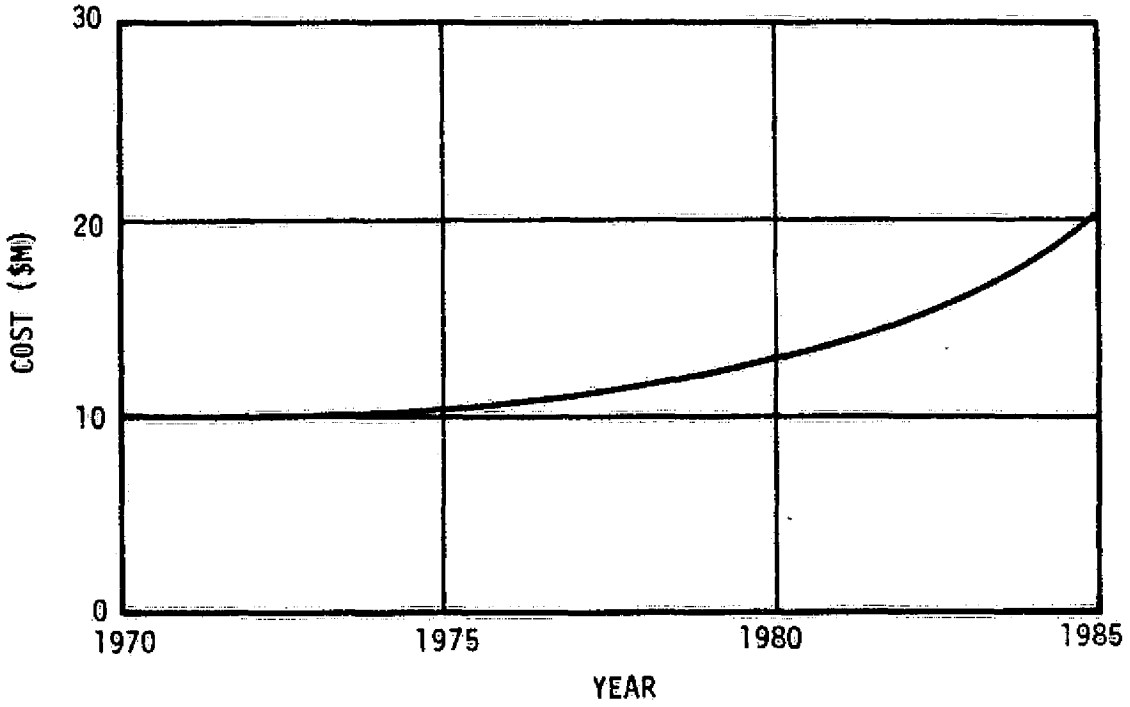


FIGURE 7.1.4.2-4. ARRAY SUPER-SCALE COMPUTER COST TREND

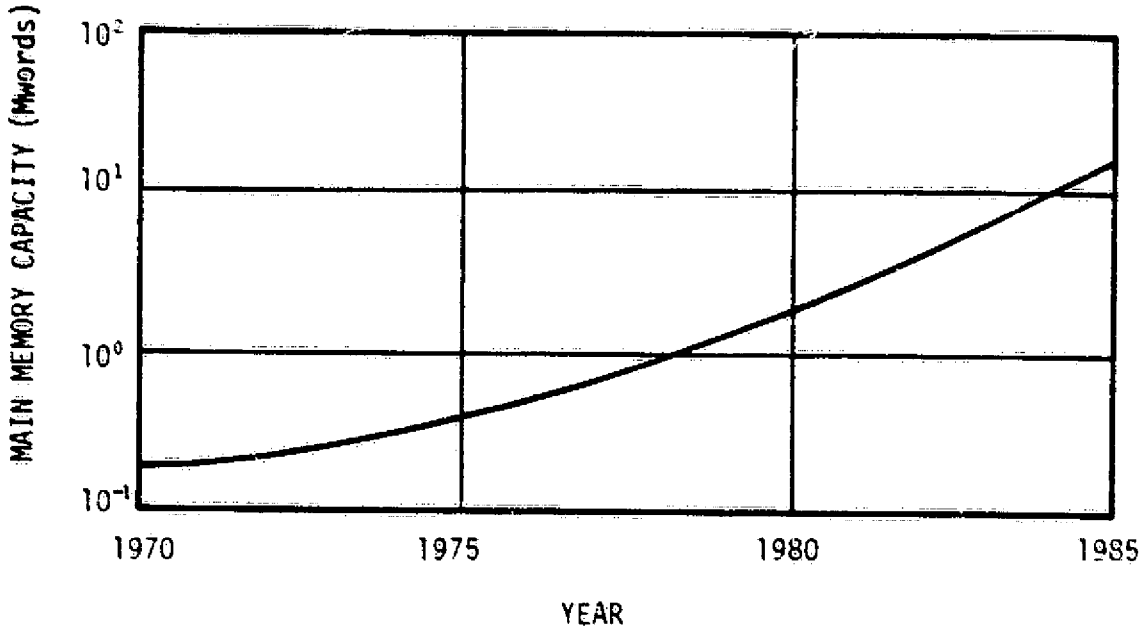


FIGURE 7.1.4.2-5. VECTOR SUPER-SCALE COMPUTER MAIN MEMORY CAPACITY TREND

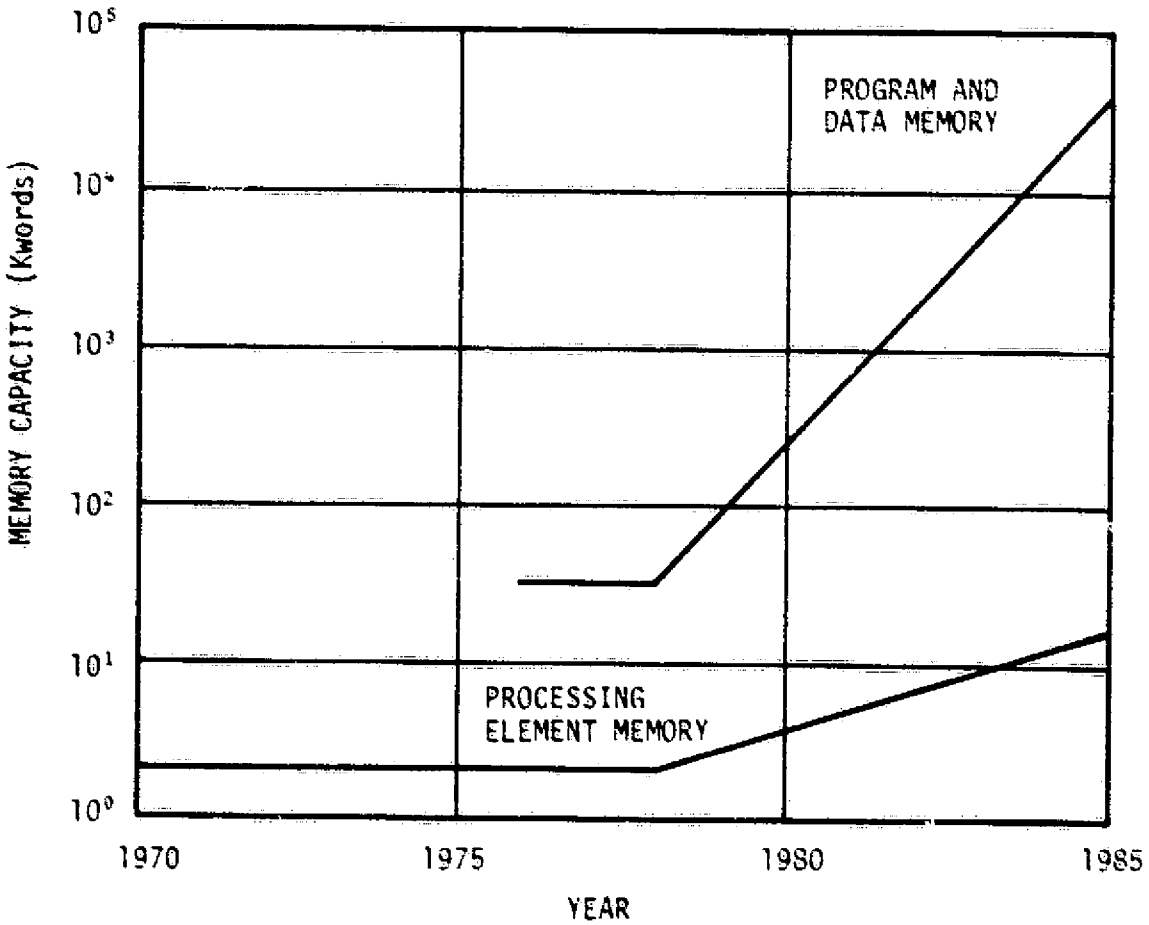


FIGURE 7.1.4.2-6. ARRAY SUPER-SCALE COMPUTER MEMORY CAPACITY TREND

logic and memory speeds will be realized through the use of these denser chips, yielding improved system performance. Array processor systems will include larger numbers of processing elements to further increase system performance.

In the 1980's, emphasis on system reliability will increase. More sophisticated error correcting and detecting codes will be used. There will be a continuation of on-line monitoring and diagnostic capabilities with automatic re-try.

Major advances are needed in software support for super-scale computers. Problem programming for a parallel or vector machine must be made reasonably straightforward and simple so that the system can be more fully utilized. For advances to occur, a better understanding of the hardware of the super-scale computers must be developed. Emphasis will be placed on the development of this understanding, and of software.

7.1.4.3 Projected Developments in Super-Scale Computers - Improvements in LSI logic circuits and semiconductor memories will continue to occur through 1985 (see Subsection 7.2.1). These improvements will be reflected in improvements in super-scale computers in many areas, including:

- Processing speed
- Memory capacity and speed
- Reliability
- Price/performance.

Tables 7.1.4.3-1 and 7.1.4.3-2 summarize the projected characteristics of vector super-scale computers in the early and mid 1980's, respectively. Table 7.1.4.3-3 summarizes the projected characteristics of array super-scale computers in the 1980-1985 timeframe. The instruction execution rates presented in these tables are conservative projections. The rates that will be achieved in super-scale computers of the next decade will depend on two things: 1) the extent to which the parallel or vector capabilities are increased, which depends primarily on hardware advances, and 2) the extent to which the parallel or vector capabilities are utilized, which depends primarily on software advances. Since software development is slower than hardware development, it will be the limiting factor, preventing the more optimistic projections made by some experts from being achieved. The basic architecture of these machines will be the same as that of current machines.

Super-scale computers will continue to have a very limited market among installations with specialized applications that are highly parallel in nature. These applications require an ever-increasing number of computations to be performed on an ever-increasing amount of data. To meet these requirements, ECL and CML logic circuits and ECL memories with large capacities will be included in the super-scale computers of the next decade. System costs will be higher, since performance improvements will more than balance any lowering of component cost. The typical computer installation will not be able to cost-justify the use of a super-scale computer.

TABLE 7.1.4.3-i. PROJECTED TECHNOLOGY FOR VECTOR SUPER-SCALE COMPUTERS, EARLY 1980's

CHARACTERISTICS	
CPU	
Word Length (bits)	64
Type of Logic Circuitry	ECL, CML, possibly GaAs
Cycle Time (nsec)	10 to 30
Minimum Computation Times:	
64-bit Add (nsec)	10 to 15
64-bit Multiply (nsec)	10 to 35
64-bit Divide (nsec)	30 to 60
Maximum Vector Result Rate (nsec)	10 to 30
Number of Instructions:	
Scalar	NA
Vector	NA
Total Number of Registers	Hundreds to 1,000
MIPS	100 to 150
Memory	
Type	TTL or ECL
Cycle Time (nsec)	100 to 35
Maximum Capacity (words)	2.5M
Bandwidth (Mwords/sec)	NA
Error Correction	Single error correcting
Virtual Memory	A few systems
I/O Maximum Rate (bytes/sec)	NA
Software	
Assemblers	Macroassembler
High-Level Languages	FORTRAN
Monitors/Operating Systems	Batch Operating

TABLE 7.1.4.3-1 - Concluded

CHARACTERISTICS	
Physical Characteristics	
Volume	NA
Mainframe Weight (tons)	NA
Cost ($M)	12.5
MTBF (hr)	100
Maximum Number of Pipelines	NA

TABLE 7.1.4.3-2. PROJECTED TECHNOLOGY FOR VECTOR SUPER-SCALE COMPUTERS, MID 1980's

CHARACTERISTICS	
CPU	
Word Length (bits)	64
Type of Logic Circuitry	GaAs or Short-Channel MOS
Cycle Time (nsec)	2 to 10
Minimum Computation Times:	
64-bit Add (nsec)	2 to 10
64-bit Multiply (nsec)	2 to 15
64-bit Divide (nsec)	10 to 30
Maximum Vector Result Rate (nsec)	2 to 10
Number of Instructions:	
Scalar	NA
Vector	NA
Total Number of Registers	NA
MIPS	200 to 300
Memory	
Type	ECL
Cycle Time (nsec)	15
Maximum Capacity (words)	16M
Bandwidth (Mwords/sec)	NA
Error Correction	Double error correcting
Virtual Memory	A few systems
I/O Maximum Rate (bytes/sec)	NA
Software	
Assemblers	Macroassembler
High-Level Languages	FORTRAN, etc.
Monitors/Operating Systems	NA

TABLE 7.1.4.3-2 - Concluded

CHARACTERISTICS	
Physical Characteristics	
Volume (ft^3)	NA
Mainframe Weight (tons)	NA
Cost ($M)	15
MTBF (hr)	200
Maximum Number of Pipelines	NA

TABLE 7.1.4.3-3. PROJECTED TECHNOLOGY FOR ARRAY SUPER-SCALE COMPUTERS, 1980-1985 TIMEFRAME

CHARACTERISTICS	
Processing Elements	
Number per System	Up to 1,000
Word Length (bits)	64
Type of Logic Circuitry	CML, GaAs or Short-Channel MOS
Cycle Time (nsec)	30
Processing Element Data Memory	
Type	ECL
Cycle Time (nsec)	50
Maximum Capacity (words)	16K
Error Correction	Yes
MIPS	NA
System MIPS	500 to 1,000
Program and Data Memory	
Type	NMOS
Cycle Time (nsec)	80 to 120
Maximum Capacity (words)	40M
Error Correction	Yes
I/O Maximum Rate	NA
Software	Minimal
System Cost ($M)	20
System Up-Time (hr/5-day wk)	100

7.2 DATA STORAGE ELEMENTS

Data storage elements are discussed in terms of two major categories considered to be representative of memory technology in general. The two categories, illustrated in Figure 7-3, are moving media memory and electronic/optical memory. The electronic/optical memory category encompasses semiconductor, core, and other electronic/optical memory and is covered in Subsections 7.2.1 through 7.2.3. The moving media category encompasses disk and tape storage and is covered in Subsections 7.2.4 and 7.2.5.

Many of the memory types covered in the subsections below are impacting other memory types in ways that are not always obvious. One of the impacts that is obvious is that which MOS RAM is having on core. For several years, the high speed and low cost of MOS memories has been driving core out of all applications except those that require long-term primary storage memory retention after loss of primary power. Although currently well-developed MOS technologies are volatile, short-term retention of data is easily provided by capacitor or battery backup of the primary power source. Long-term storage after power failure may be provided by "rolling out" the data to some nonvolatile secondary storage (using backup power to complete the transfer) and "rolling in" the data after primary power is restored.

Less obvious is the impact MOS RAMs and ROMs will soon have on bipolar RAMs and ROMs. In theory, MOS devices should be faster than bipolar devices due to MOS's use of majority rather than minority carriers. Until recently, this advantage had never been realized in practice; however, the short-channel MOS devices make higher-speed MOS technology a reality. In addition to the short-channel MOS types, conventional MOS memories have been gaining on bipolar memories as device geometries are scaled down.

Several types of electronic/optical memories are (or will be) impacting moving-media storage devices. CCDs and magnetic bubbles are just beginning to invade the markets currently served by magnetic disk

and drum. (Drum seems to be on the way out anyway, and will not be considered further.) Both CCDs and bubbles promise to equal or exceed the performance of disk devices in all areas. CCD has a speed advantage over bubbles, but bubbles are nonvolatile. Thus each may have a place in the marketplace as electronic replacements for disks. Disks will probably retain a significant share of the market, but only in those applications requiring removable media (e.g., small computers using disk for changeable file storage, etc.).

Although not currently important, laser/holographic storage shows some promise of impacting reel-to-reel magnetic tape where the latter is used for archival storage. The high packing density and ease of storage (no periodic rewinding of tapes to retard printthrough) achievable with laser/holographic means is attractive, but actual impact will depend on how fast these systems are developed.

7.2.1 Semiconductor Memory

In addition to the technologies covered in Subsections 7.2.1.1 through 7.2.1.3, several new semiconductor technologies are emerging that have possible application in semiconductor memories. The general aspects of these new technologies are covered in Section 1 of this report. The discussion here is limited to those aspects of these technologies that directly relate to memory applications.

Several new device types, such as Static Induction Transistors (SITs) and Josephson junctions, and new semiconductor materials, such as gallium-arsenide (GaAs), are impacting semiconductor memories. SITs are a type of vertical field-effect transistor (VFET) currently receiving much attention in research. In memory applications, work has been done that suggests that RAMs are possible with very high density (>64 Kbits), high speed (subnanosecond switching, few-nanosecond access), and low power (a few microwatts per bit). Research into these memories is currently at an early stage, so the future of these devices in memory applications is unclear.

Research on using Josephson junction devices in RAMs is also underway. Current work is aimed at proving the feasibility of 16-Kbit, 15-nsec access time, 30-nsec cycle time memories using Josephson junction devices as the storage cell. Other projections for the Josephson junction technology being developed by IBM include a fanout limited by permissible delay (3.4 psec/pH inductance of the output line), a speed-power product of 6×10^{-7} J/pH of output inductance, a power dissipation of 10 μW/bit if the chip is selected, and zero power dissipation in a nonselected chip.

In the area of new semiconductor material, GaAs is the type receiving the most attention. Most of the research into digital devices using GaAs as the base semiconductor has been in the area of logic circuits rather than memory circuits. The same benefits that use of GaAs gives in logic circuits should be realizable in memory circuits: i.e., high speed and good speed-power product.

7.2.1.1 Bipolar Memory -

7.2.1.1.1 State of the Art in Bipolar Memory - TTL, ECL, and I^2L technologies are used to build bipolar memory chips. For each of these three technologies, Table 7.2.1.1.1-1 presents a summary of the characteristics of state-of-the-art RAM chips and of the memory systems built from these chips. TTL and ECL RAM chips are used to build memories for high-performance applications, such as scratchpad and buffer memories, control stores, and fast main memories, where operating speed is very important. ECL is the fastest bipolar technology but has the highest power dissipation. In general, for both TTL and ECL chips and memory systems, the fastest access and cycle times listed in Table 7.2.1.1.1-1 are associated with the highest cost per bit and the largest power dissipation. The packing of 1 to 4 Kbits on TTL and ECL chips is opening the way for practical use of bipolar RAMs in larger systems because the ease of interfacing such RAMs allows the design of memory systems in which control and peripheral circuitry add little to the system's total access time. The 4-Kbit TTL and ECL chips listed in Table 7.2.1.1.1-1 are just reaching production.

Integration injection logic (I^2L) is the newest bipolar technology. It is being used to design dynamic bipolar memory chips with potentially more bits per chip than achieved using either TTL or ECL technology. I^2L is believed to have the density and low-cost advantage of the MOS process as well as speeds only slightly lower than those achieved with TTL technology. As in the case of the other bipolar technologies, the fastest access and cycle times listed in Table 7.2.1.1.1-1 are associated with the highest cost per bit and the largest power dissipation. Nearly every semiconductor manufacturer is either actively investigating or currently introducing I^2L memory products.

TABLE 7.2.1.1.1-1. STATE OF THE ART IN BIPOLAR MEMORIES

CHARACTERISTICS	TTL	4K TTL	ECL	4K ECL	I^2L
Chip					
Capacity (Kbits)	1	4	1	4	4
Size (mil^2)	15,000	NA	15,000	23,600	14,500
Maximum Access Time (nsec)	30 to 60	50 to 70	10 to 35	25 to 40	50 to 100
Minimum Cycle Time (nsec)	50 to 100	70 to 120	25 to 50	40 to 60	100 to 200
Maximum Power Dissipation (μW/bit)	500 to 200	400 to 300	812.5 to 500	400	122 to 12.2
Cost (¢/bit)	1.5 to 0.6	NA	1.5 to 0.6	NA	0.7 to 0.2
Type	Static	Static	Static	Static	Dynamic
MTBF (hr)	10^7 to 10^8	NA	10^7 to 10^8	NA	NA
System					
Capacity (bits)	Up to 64M	Up to 128M	Up to 64M	Up to 128M	Up to 128M
Physical Volume	NA	NA	NA	NA	NA
Maximum Access Time (nsec)	60 to 100	80 to 120	25 to 55	40 to 60	80 to 140
Minimum Cycle Time (nsec)	80 to 140	100 to 160	32 to 70	47 to 80	130 to 240
Maximum Power Dissipation (μW/bit)	512 to 212	412 to 312	836 to 524	420	134 to 24
Cost (¢/bit)	3.40 to 1.20	NA	3.40 to 1.20	NA	0.97 to 0.31
MTBF (hr) (with error correction on larger memories)	10^3 to 10^4	NA	10^3 to 10^4	NA	NA

7.2.1.1.2 Trends in Bipolar Memory - Many competent organizations are developing and producing bipolar RAM chips in a highly competitive market. This will assure continued improvements in all aspects of the three bipolar memory technologies. Chips will get both larger and denser, as illustrated in Figure 7.2.1.1.2-1 and Table 7.2.1.1.2-1, and the growth rate for chip density will continue to exceed the growth rate for physical chip size. As the chips become denser, chip and memory system speeds will increase, as illustrated in Figure 7.2.1.1.2-2. Improvements in manufacturing processes and yields should continue to result in cost per bit reductions, as illustrated in Figure 7.2.1.1.2-3. System-level cost per bit will decrease more rapidly than chip-level cost per bit because when denser chips are used in a memory system the memory chip cost is a greater percentage of the total system cost. Although power dissipation per chip will increase slowly, power dissipation per bit should decline steadily, as illustrated in Figure 7.2.1.1.2-4. This trend will aid the development of denser chips because power consumption and removal of the heat generated by it are limiting factors to the growth in bipolar chip density.

As a more mature technology, bipolar RAMs, especially TTL and ECL, will improve at a slower rate than MOS RAMs (see Subsection 7.2.1.2.2). TTL and ECL RAMs will continue to require about five times the chip area of MOS RAMs because of the complexity of circuitry and the need for passive components, to cost two to five times as much per bit, and to dissipate substantially more power. I^2L RAMs will suffer much less from these handicaps.

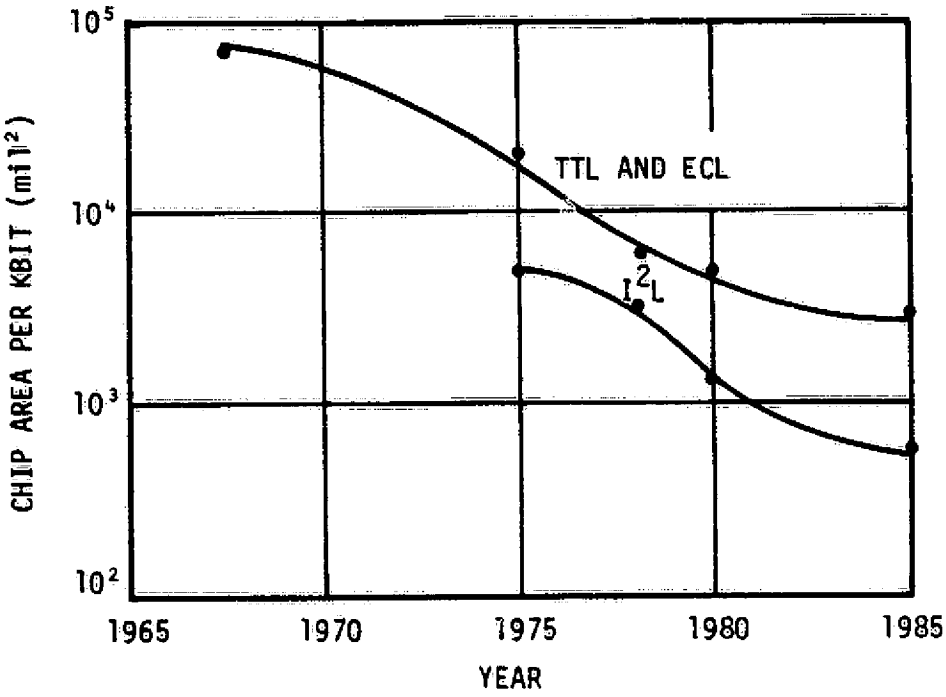


FIGURE 7.2.1.1.2-1. BIPOLAR MEMORY CHIP SIZE TRENDS

TABLE 7.2.1.1.2-1. BIPOLAR MEMORY CHIP DENSITY TRENDS

YEAR	TTL	ECL	I^2L
1965	16		
1966			
1967			
1968			
1969	64		
1970		256	
1971			
1972			
1973			
1974	1K		
1975		1K	
1976			
1977			4K
1978	4K	4K	16K
1979			
1980			
1981			
1982			64K
1983	16K		
1984		16K	
1985			

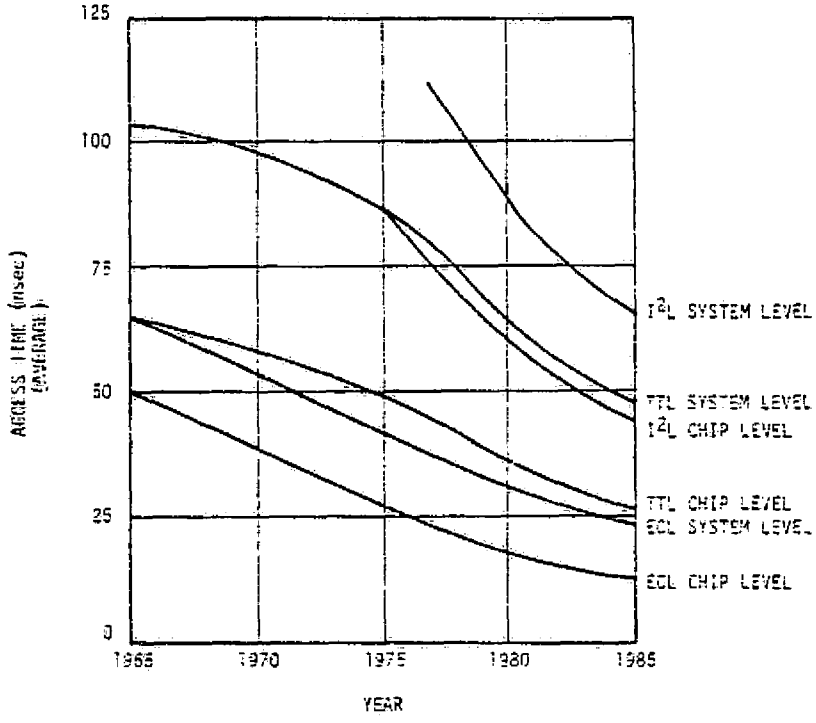


FIGURE 7.2.1.1.2-2. BIPOLAR MEMORY ACCESS TIME TRENDS

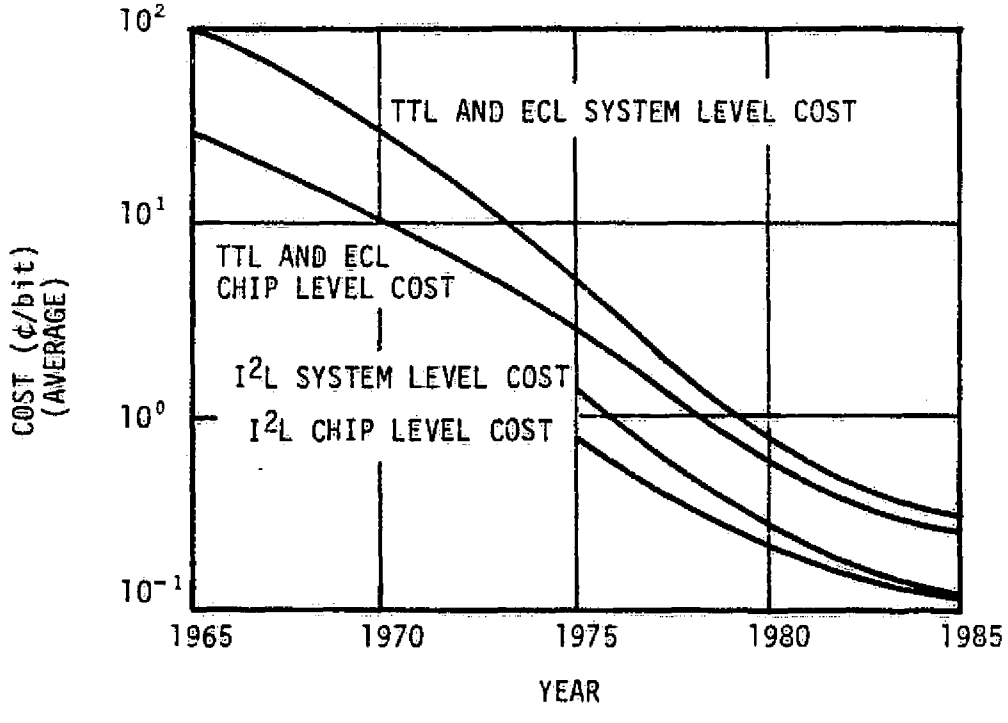


FIGURE 7.2.1.1.2-3. BIPOLAR MEMORY COST TRENDS

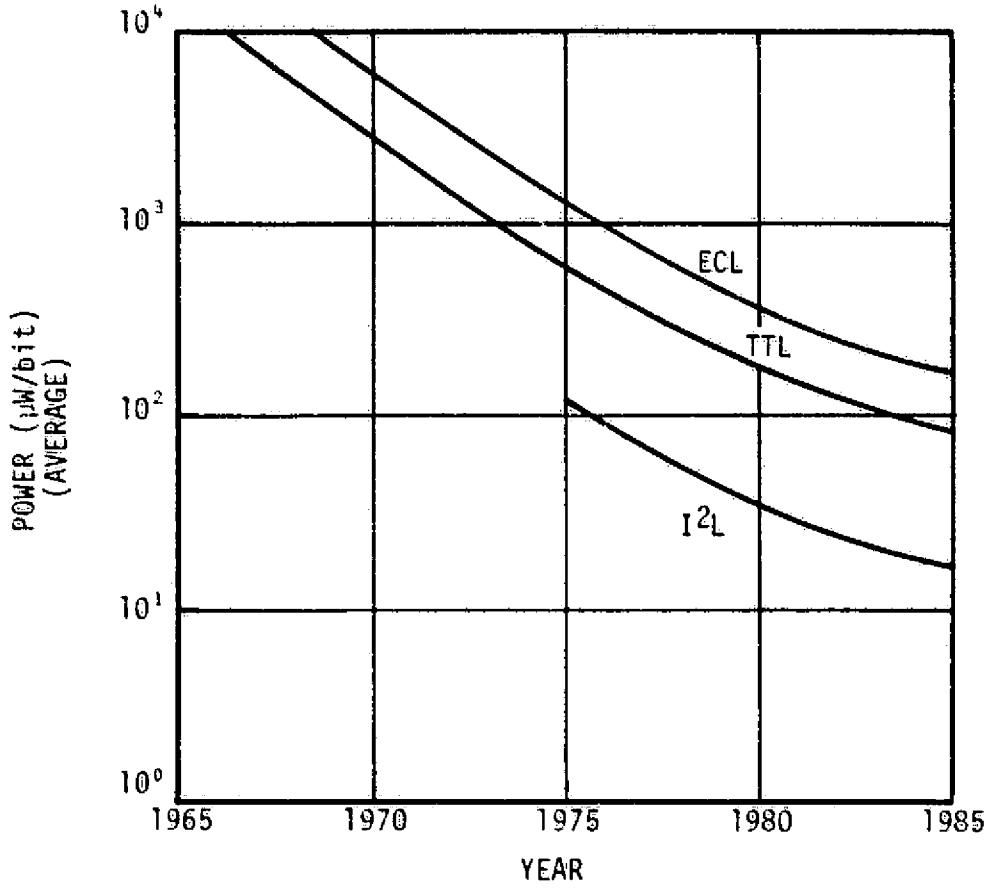


FIGURE 7.2.1.1.2-4. BIPOLAR MEMORY POWER DISSIPATION TRENDS

7.2.1.1.3 Projected Developments in Bipolar Memory - The trends in bipolar technology over the past decade and the assessment of several authorities in the field indicate that bipolar RAMs will be commercially available in 1980 and 1985 with the characteristics given in Tables 7.2.1.1.3-1 and 7.2.1.1.3-2, respectively. As with current RAM chips and memory systems, the fastest access and cycle times listed in these tables are associated with the highest cost per bit and the largest power dissipation. The 1980 and 1985 system-level access times, cycle times, and power dissipation figures were obtained by assuming that the same overhead circuitry will be needed in future systems that is needed in current systems and that improvements in the characteristics of this circuitry will occur at the same rate as improvements in these same characteristics on the chips themselves.

TTL and ECL RAMs will continue to be used to meet fast-access requirements, appearing both as buffer memory and as main memory when the cost can be justified. The fastest bipolar semiconductor memories can be expected to use ECL or some circuitry to be developed. The newness of I^2L technology leaves a greater degree of uncertainty as to what its role will be in memory systems during the next decade. One group of experts believes that I^2L RAMs will be the next generation of high-density, low-cost components for main memories, competing with NMOS RAMs. Another group sees more use for I^2L RAMs and ROMs as elements for Programmable Logic Arrays (PLAs). As a new technology, I^2L is in need of refinements to achieve better production yields and higher speeds.

Improvements in chip-level reliability will occur during the next decade, with resultant increases in system-level reliability. However, the chips will not reach the degree of reliability required to enable large bipolar memory systems to reach the desired level of system reliability. To achieve high system reliability for these large memory systems, error correcting codes will be used.

TABLE 7.2.1.1.3-1 PROJECTED TECHNOLOGY IN BIPOLAR MEMORY, 1980

CHARACTERISTICS	TTL	ECL	I^2L
Chip			
Capacity (Kbits)	4	4	16
Size (mil^2)	20,000	20,000	20,000
Maximum Access Time (nsec)	24 to 48	7 to 28	40 to 80
Minimum Cycle Time (nsec)	40 to 80	19 to 40	80 to 160
Maximum Power Dissipation (μW/bit)	250 to 100	406 to 250	61 to 6.2
Cost (¢/bit)	0.75 to 0.3	0.75 to 0.3	0.35 to 0.1
Type	Static	Static	Dynamic
MTBF (hr)	10^7 to 10^8	10^7 to 10^8	NA
System			
Capacity (bits)	Up to 160M	Up to 160M	Up to 492M
Physical Volume	NA	NA	NA
Maximum Access Time (nsec)	48 to 80	19 to 44	64 to 112
Minimum Cycle Time (nsec)	64 to 112	25 to 56	104 to 192
Maximum Power Dissipation (μW/bit)	256 to 106	418 to 262	67 to 12
Cost (¢/bit)	1.0 to 0.41	1.0 to 0.41	0.39 to 0.12
MTBF (hr) (with error correction on larger memories)	10^4	10^4	NA

TABLE 7.2.1.1.3-2. PROJECTED TECHNOLOGY IN BIPOLAR MEMORY, 1985

CHARACTERISTICS	TTL	ECL	I^2L
Chip			
Capacity (Kbits)	16	16	64
Size (mil^2)	44,000	44,000	38,000
Maximum Access Time (nsec)	18 to 36	5 to 21	30 to 60
Minimum Cycle Time (nsec)	30 to 60	14 to 30	60 to 120
Maximum Power Dissipation (μW/bit)	125 to 50	203 to 125	30 to 3
Cost (¢/bit)	0.375 to 0.15	0.375 to 0.15	0.175 to 0.05
Type	Static	Static	Dynamic
MTBF (hr)	10^8 to 10^9	10^8 to 10^9	NA
System			
Capacity (bits)	Up to 1G	Up to 1G	Up to 2G
Physical Volume	NA	NA	NA
Maximum Access Time (nsec)	36 to 60	14 to 33	48 to 84
Minimum Cycle Time (nsec)	48 to 84	19 to 42	78 to 144
Maximum Power Dissipation (μW/bit)	128 to 53	209 to 131	33 to 6
Cost (¢/bit)	0.45 to 0.2	0.45 to 0.2	0.19 to 0.06
MTBF (hr) (with error correction on larger memories)	10^5	10^5	NA

7.2.1.2 MOS Memory -

7.2.1.2.1 State of the Art in MOS Memory - PMOS, NMOS, CMOS, MNOS, VMOS, and DMOS technologies are currently used to build MOS memory chips. Table 7.2.1.2.1-1 presents a summary of the characteristics of state-of-the-art RAM chips and of the memory systems built from these chips. For each of the chips and systems summarized, the fastest access and cycle times listed in this table are associated, in general, with the highest cost per bit and the largest power dissipation. Reduced cost per bit and reduced overhead circuit requirements have made semiconductor RAMs attractive for the replacement of magnetic core memories in computer mainframes. For most new developments, semiconductor RAMs are dominant up to 10^6 bits, which is 64K words in a 16-bit minicomputer system, and semiconductor memory systems have been built containing more than 10^8 bits. Currently, about 60% of the new main memory market belongs to semiconductor memories, with 40% belonging to MOS memories and 20% belonging to bipolar memories. MOS memories already play a strong role in large-scale memory systems. Although core technology still dominates the installed base of minicomputers, MOS technology dominates the new announcements. The principal advantages of MOS memories are higher speed, lower cost, reduced space requirements, lower power consumption, nondestructive readout, and simplified maintenance. The main disadvantage is volatility, and the newer MNOS technology is designed to offer nonvolatile semiconductor storage.

PMOS is the oldest MOS technology. The 4-Kbit memory chip represents the state of the art for this technology. Since the newer NMOS technology offers nearly a factor of two increase in operating speeds because of lower capacitances and lower threshold voltage, PMOS is basically an obsolete technology.

NMOS is the MOS technology that offers the best combination of memory access time, chip density, power dissipation, and cost per bit. Both static and dynamic NMOS RAM chips are in widespread use today. Dynamic chips store data as a capacitive charge and periodically refresh

TABLE 7.2.1.2.1-1. STATE OF THE ART IN MOS MEMORY

CHARACTERISTICS	DMOS	NMOS	VMOS	NMOS	CMOS	MNOS
Chip						
Capacity (Kbits)	4	4	4	16	1	4
Size (mil^2)	33,000	30,000	14,000	30,000	11,200	37,000
Maximum Access Time (nsec)	60	45 to 250	55 to 150	150 to 250	60 to 300	1,600 (read) 100 msec (write)
Minimum Cycle Time (nsec)	180	130 to 320	NA	375 to 410	450 to 600	NA
Maximum Power Dissipation (μW/bit)	232	160 to 90 active 160 to 5 standby	120 to 110 active 25 standby	45 to 30 active 1.2 to 0.6 standby	20 to 10 active 3 to 0.05 standby	75 to 100
Cost (¢/bit)	NA	0.35 to 0.15	0.61	0.3 to 0.1	2.8 to 0.9	0.3
Type	Dynamic	Static	Static	Dynamic	Static	Static
MTBF (hr)	NA	10^7 to 10^8	10^6 to 10^7	10^7 to 10^8	NA	NA
System						
Capacity (bits)	Up to 128M	Up to 128M	Up to 128M	Up to 128M	NA	Up to 16M*
Physical Volume	NA	NA	NA	NA	NA	NA
Maximum Access Time (nsec)	90 to 100	85 to 290	90 to 190	190 to 290	100 to 350	2,000 per 1K block
Minimum Cycle Time (nsec)	210	160 to 450	NA	400 to 450	480 to 640	4,000 per 1K block (read) 200 msec per 1K block (write)
Maximum Power Dissipation (μW/bit)	256	172 to 102 active 172 to 10 standby	134 to 122 active 134 to 33 standby	57 to 42 active 7 to 5 standby	32 to 22 active 8 to 5 standby	NA
Cost (¢/bit)	NA	0.50 to 0.25	0.75	0.34 to 0.12	4.60 to 1.60	NA
MTBF (hr)	NA	10^3 to 10^4	NA	10^3 to 10^4	NA	NA

*BORAM (Block Organized RAM) systems

this charge before it can leak away, as opposed to static chips that store data in a bistable multivibrator and do not require periodic refresh. A dynamic cell requires only a fraction of the transistors needed by a static cell and also does not require a dc power component, although the sense amplifier may or may not dissipate dc power depending on the particular memory system design. The ac power dissipation of a dynamic cell is a function of the rate at which the memory is cycled, increasing as the cycle time increases. Obviously, density is greatly improved with a dynamic chip and power requirements are potentially much less. Dynamic chips are also less expensive and faster. However, they do require constant memory refreshing from additional support circuitry.

For small systems that cannot tolerate the additional support circuitry required by dynamic RAMs and for systems that cannot allocate time for memory refresh, static RAMs are essential. The state of the art for static NMOS RAMs is the 4-Kbit chip, organized as either 4K × 1 bits or 1K × 4 bits, packaged in an 18-pin dual in-line package, and requiring a single +5-V supply. Chip and system characteristics are listed in Table 7.2.1.2.1-1. A few of the high-performance chips still use multiple power supplies. Static RAM chips are available in both static and quasi-static configurations. Both configurations employ a static memory array, but in the quasi-static configuration, power applied to peripheral circuits can be switched off to conserve power, allowing standby power dissipation to be much lower than active power dissipation at the expense of circuit complexity. For the static configuration, standby and active power dissipation are the same, but this RAM chip is simpler to design into a memory system.

For main memory applications, one is primarily concerned with dynamic RAMs because of the power requirements of static RAMs. The state of the art for dynamic NMOS RAMs is the 4-Kbit and 16-Kbit chips, with the 16-Kbit chips now reaching full production status. Several companies are known to have prototype or planned 64-Kbit dynamic NMOS RAMs, including Nipon Telegraph and Telephone, Texas Instruments, and

Mostek. Sample devices are anticipated for mid-1978, with full production due in 1979. Chip and system characteristics for 16-Kbit RAMs are listed in Table 7.2.1.2.1-1. The large-board, dense RAM chip systems cost substantially less than their smaller counterparts because both the storage and the nonstorage costs per bit decrease with increasing chip density.

CMOS is the high-speed, low-power-dissipation, high-noise-immunity MOS technology. Its state of the art is the 1-Kbit chip with the characteristics listed in Table 7.2.1.2.1-1. CMOS memory system characteristics are also listed in this table. CMOS chips utilizing SOS (silicon on sapphire) technology have the fastest access and cycle time but also cost the most per bit. Chips not using the SOS technology are available for under 1 ¢/bit but have access times greater than 100 nsec. Because of the very low power requirements, 1-Kbit CMOS RAMs that operate from a small battery are available, thus providing a nonvolatile, portable memory package. Since CMOS is a static design that uses more silicon per cell than other technologies, it is generally agreed that the high cost of silicon will make CMOS useful for memories no larger than 1 Kbit in capacity, except for special applications for which some 4-Kbit CMOS chips are available.

MNOS is the MOS technology that attempts to overcome the volatility of semiconductor memories through the use of a double insulator layer structure. Its state of the art is the 4-Kbit chip. Chip and system characteristics are listed in Table 7.2.1.2.1-1. Expected data-retention time quoted for state-of-the-art MNOS memory systems is 10 years, and researchers claim to have increased this time to 200 years in laboratory models. However, the long-term reliability of MNOS memory systems is still greatly debated. Because of the large access and write times, much improvement is needed for MNOS memories to become a viable alternative in general-purpose memory applications.

VMOS and DMOS are new MOS technologies that increase speed by shortening the effective channel length of the individual MOS devices.

Their state of the art is the 4-Kbit chip, with some manufacturers known to be working on 64-Kbit VMOS chips. Chip and system characteristics for 4-Kbit devices are listed in Table 7.2.1.2.1-1. The potential speed advantage of MOS devices over bipolar devices (due to MOS's use of majority carriers rather than minority carriers as for bipolar) is enhanced for VMOS and DMOS devices. This is because the short channel length of these devices lowers the gate capacitance, allowing it to charge more quickly, and thus speeding up the switching of the device. Power dissipation per bit for both VMOS and DMOS is currently higher than for other MOS technologies, but should come down dramatically with further research and increasing chip densities. DMOS is also known as DSAMOS (Diffusion Self-Aligned MOS). Development of DMOS is going on mostly in Japan, while VMOS is being developed mostly by domestic companies.

7.2.1.2.2 Trends in MOS Memory - Many competent organizations are developing and producing MOS memory chips in a highly competitive market. This will assure continued significant improvements in all aspects of MOS technology and will secure the position for MOS memories as the dominant processor memory technology, at least through 1985.

NMOS RAM chips will get both larger and denser, as illustrated in Figure 7.2.1.2.2-1 and Table 7.2.1.2.2-1, and the growth rate for chip density will continue to exceed the growth rate for physical chip size. These improvements will be the result of continued improvements in lithography, fabrication, and packaging. The historic factor-of-four improvement in NMOS RAM chip density approximately every 2 years will continue during the next several years. When photolithographic limits are reached with the 256-Kbit NMOS RAM chip, X-ray lithography or direct-electron-beam etching will be available. The 256-Kbit chip will benefit from direct-electron-beam die etching and/or X-ray lithography because of the potential yield increase due to smaller chips. In the opinion of most experts, cost-effective use of these techniques is 3 to 4 years away for electron-beam etching and 5 more years away for X-ray lithography. CMOS RAM chips will probably get smaller, with the chip density maintained at 1 to 4 Kbits because of the high cost of silicon.

As the chips become denser, chip and memory system speeds will increase, as illustrated in Figures 7.2.1.2.2-2 and 7.2.1.2.2-3. Since semiconductor memories generally rely on increasing chip density to reduce cost per bit, the growth in chip density, aided by improvements in manufacturing processes and yields, will allow the trend of a 30 to 35% cost per bit reduction each year to continue (Ref. 7-11), as illustrated in Figure 7.2.1.2.2-4. System-level cost per bit will decrease more rapidly than chip level cost per bit, because when denser chips are used in a memory system, the memory chip cost is a greater percentage of the total system cost. Although power dissipation per chip will increase slowly during the next decade, power dissipation per bit should decline steadily, as illustrated in Figure 7.2.1.2.2-5. This trend will aid the development of denser chips. Reliability at the chip level will continue to increase as denser chips and improved fabrication processes are developed.

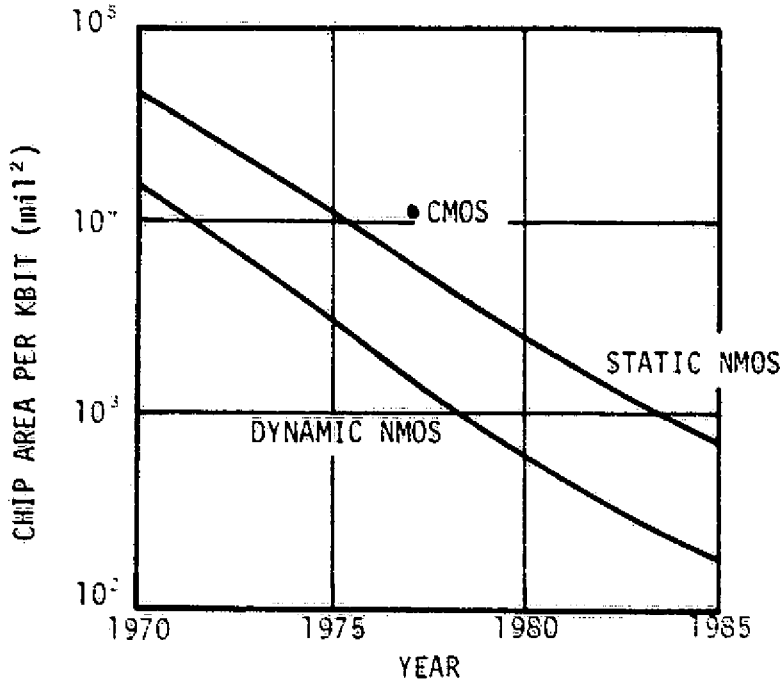


FIGURE 7.2.1.2.2-1. MOS MEMORY STORAGE DENSITY TRENDS

TABLE 7.2.1.2.2-1. MOS MEMORY CHIP DENSITY TRENDS

YEAR	STATIC	DYNAMIC
1969	256	
1970		1K
1971		
1972	1K	
1973		4K
1974		
1975		
1976	4K	16K
1977		
1978		64K
1979	16K	
1980		256K
1981		
1982	64K	1M
1983		
1984		
1985	256K*	4M*

*NOTE: These high density memory chips will use new MOS device technologies.

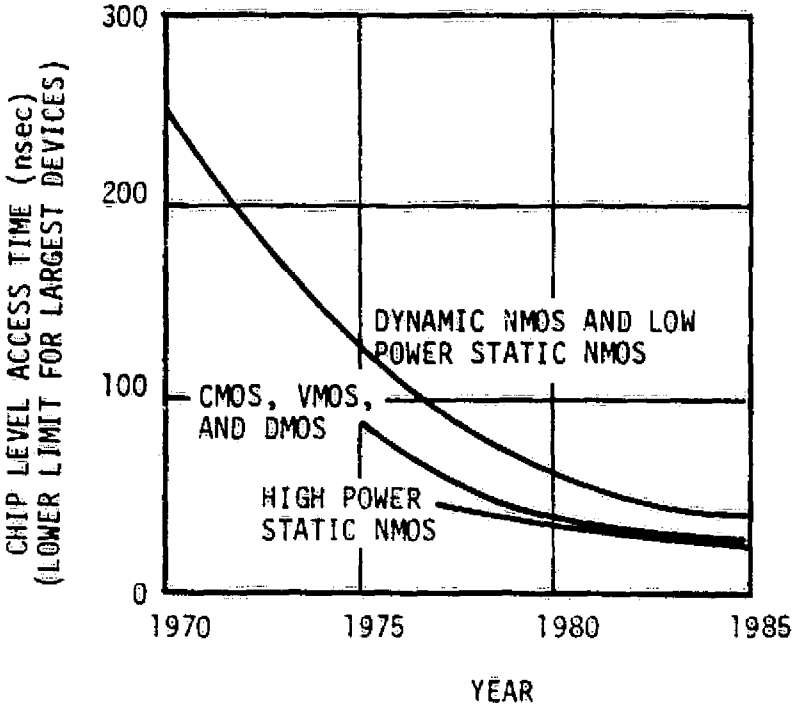


FIGURE 7.2.1.2.2-2. MOS MEMORY CHIP LEVEL ACCESS TIME TRENDS

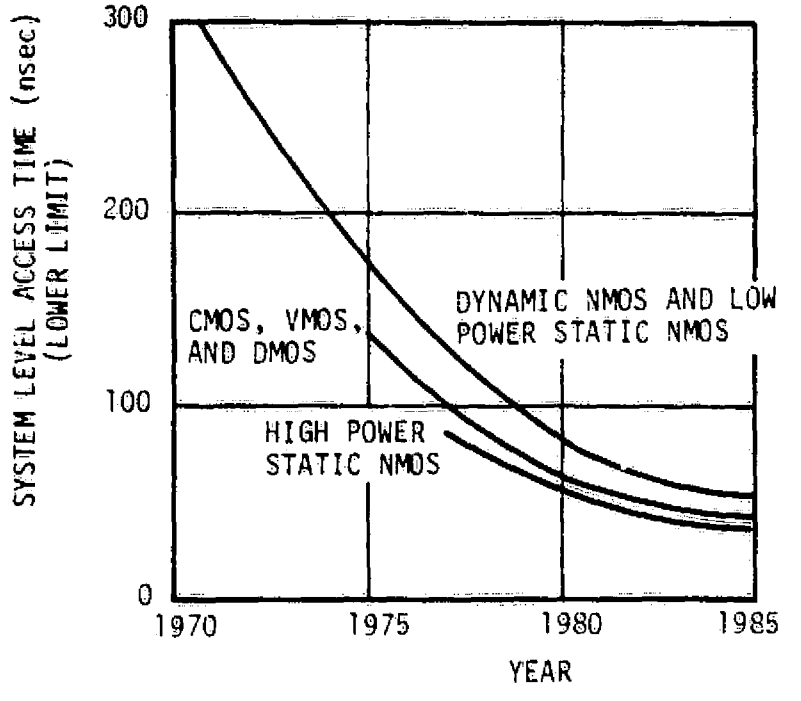


FIGURE 7.2.1.2.2-3. MOS MEMORY SYSTEM LEVEL ACCESS TIME TRENDS

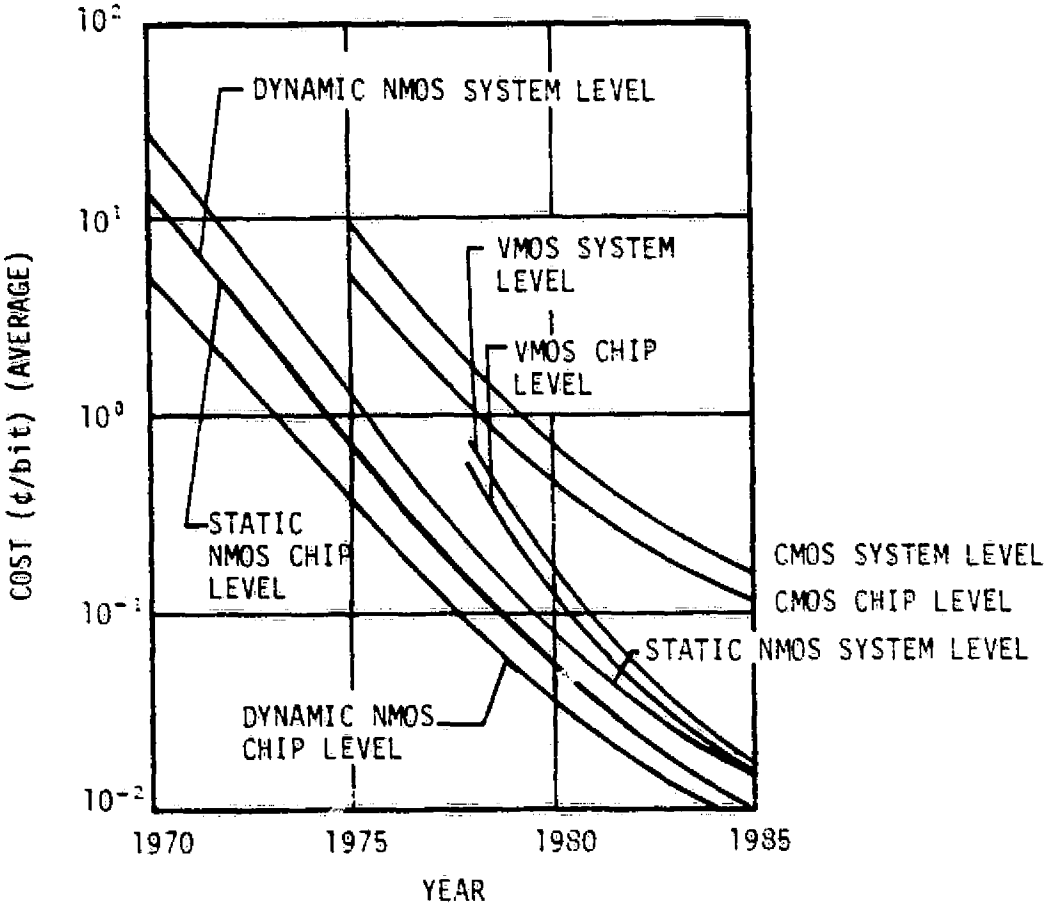


FIGURE 7.2.1.2.2-4. MOS MEMORY COST TRENDS

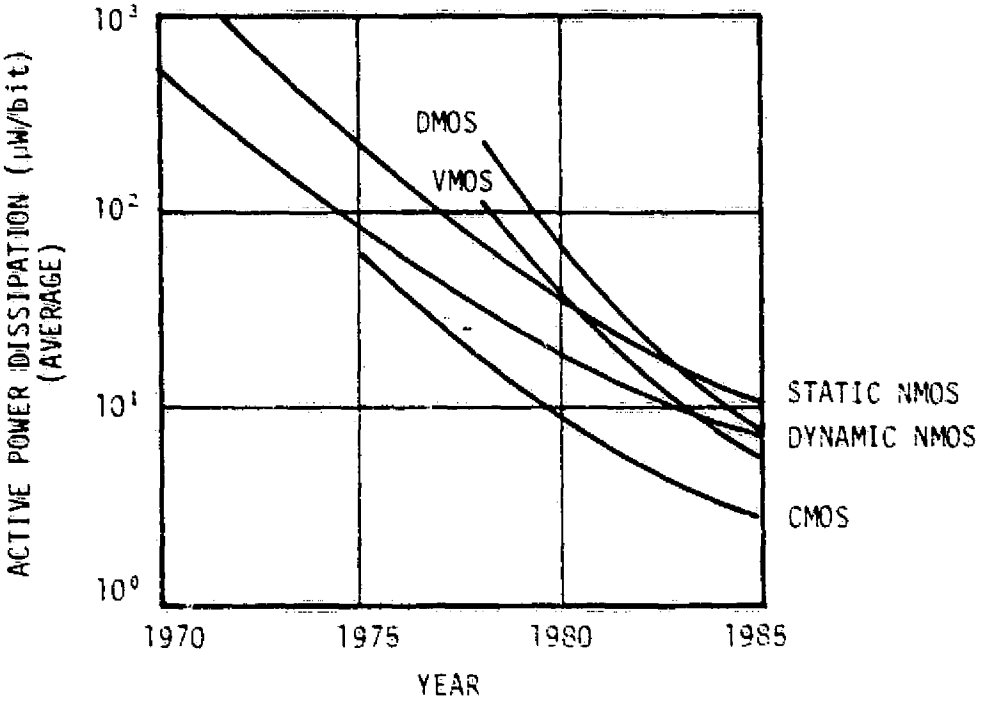


FIGURE 7.2.1.2.2-5. MOS MEMORY ACTIVE POWER DISSIPATION TRENDS

7.2.1.2.3 Projected Developments in MOS Memory - The trends in MOS technology over the past decade and the assessment of several authorities in the field indicate that NMOS and CMOS RAMs will be commercially available in 1980 and 1985 with the characteristics given in Tables 7.2.1.2.3-1 and 7.2.1.2.3-2, respectively. As with current RAM chips and memory systems, the fastest access and cycle times listed in these tables are associated with the highest cost per bit and the largest power dissipation. The 1980 and 1985 system-level access times, cycle times, and power dissipation figures were obtained by assuming that the same overhead circuitry will be needed in the future systems that is needed in current systems and that improvements in the characteristics of this circuitry will occur at the same rate as improvements in these same characteristics on the chips themselves.

MOS RAMs will be the dominant main memory technology during the next decade. However, the highest-speed devices typically will not be used in the largest-capacity systems because of their power requirements. The portable file storage area, now dominated by floppy and fixed-head disks, will be penetrated by MOS memories as the result of better MOS memory reliability and the achievement of higher densities and lower costs (Ref. 7-12). Continued improvements in fixed-head and moving-head disk and drum systems will block MOS memories from entering the mass storage area.

Improvements in chip-level reliability will occur during the next decade, with resultant increases in system-level reliability. However, the chips will not reach the degree of reliability required to enable large MOS memory systems to reach the desired level of system reliability. As MOS memories are used in larger-capacity memory systems, the use of error correcting codes will expand to achieve high system reliability. As the cost per bit continues to decline, the extra bits and associated logic required for error correcting codes become insignificant when compared with the cost of emergency service.

TABLE 7.2.1.2.3-1. PROJECTED TECHNOLOGY IN MOS MEMORY, 1980

CHARACTERISTICS		NMOS	NMOS	CMOS
Chip				
Capacity (Kbits)		16	64	1 to 4
Size (mil^2)		40,000	40,000	NA
Maximum Access Time (nsec)		35 to 150	50 to 150	36 to 180
Minimum Cycle Time (nsec)		78 to 192	90 to 219	270 to 360
Maximum Power Dissipation (μW/bit)	active	40 to 23	20 to 15	5 to 2.5
	standby	40 to 2	1.5 to 0.7	0.75 to 0.0125
Cost (¢/bit)		0.09 to 0.04	0.08 to 0.02	0.7 to 0.23
Type		Static	Dynamic	Static
MTBF (hr)		10^7 to 10^8	10^7 to 10^8	NA
System				
Capacity (bits)		Up to 512M	Up to 512M	NA
Physical Volume		NA	NA	NA
Maximum Access Time (nsec)		54 to 174	68 to 174	60 to 210
Minimum Cycle Time (nsec)		96 to 270	108 to 270	288 to 384
Maximum Power Dissipation (μW/bit)	active	46 to 29	26 to 21	11 to 8
	standby	46 to 5	5 to 4	4 to 3
Cost (¢/bit)		0.1 to 0.05	0.09 to 0.026	1.2 to 0.14
MTBF (hr)		10^4	10^4	NA

TABLE 7.2.1.2.3-2. PROJECTED TECHNOLOGY IN MOS MEMORY, 1985

CHARACTERISTICS		NMOS	NMOS	CMOS
Chip				
Capacity (Kbits)		64 to 256	1,024 to 4,096	1 to 4
Size (mil^2)		48,000	48,000	NA
Maximum Access Time (nsec)		25 to 100	35 to 100	24 to 120
Minimum Cycle time (nsec)		52 to 128	60 to 146	180 to 240
Maximum Power Dissipation (uW/bit)	active	16 to 9	9 to 7	2 to 1
	standby	16 to 0.5	0.56 to 0.28	0.3 to 0.005
Cost (¢/bit)		0.020 to 0.009	0.02 to 0.005	0.25 to 0.07
Type		Static	Dynamic	Static
MTBF (hr)		10^6 to 10^9	10^6 to 10^9	NA
System				
Capacity (bits)		Up to 26	Up to 26	NA
Physical Volume		NA	NA	NA
Maximum Access Time (nsec)		36 to 116	46 to 116	40 to 140
Minimum Cycle Time (nsec)		64 to 180	72 to 180	192 to 256
Maximum Power Dissipation (uW/Bit)	active	19 to 12	12 to 10	5 to 4
	standby	19 to 2	2	2 to 1.5
Cost (¢/bit)		0.030 to 0.012	0.03 to 0.009	0.4 to 0.09
MTBF (hr)		10^5	10^5	NA

7.2.1.3 Charge-Coupled Devices - Charge-coupled devices (CCDs) were invented by Boyle and Smith in 1969 and first described in the April 1970 Bell System Technical Journal (Ref. 7-13). Charge-coupling is a principle in semiconductor electronics whereby mobile "packets" of stored electric charge are manipulated by the application of external voltages. CCD devices are manufactured using N-channel M S (NMOS) processing technology. However, the CCD manufacturing process is somewhat simpler than for standard NMOS memories since the CCDs are contact free in the main array areas, which means that they can be processed with higher yield.

The major applications of CCD devices to date include: image sensors, sampled-signal processing devices, and memory devices. The application of CCDs to signal-processing is covered in Section 2. Recent advances in CCD technology have produced new kinds of digital logic devices and a new random-access memory technology. CCD memory technology is discussed in the subsequent subsections.

7.2.1.3.1 State of the Art in CCD Memory - CCD memory chips are currently available in 64-Kbit/chip packages. These 64-Kbit devices were first introduced in 1976 and are the highest density semiconductor memory devices currently available. State-of-the-art CCD memory device characteristics are presented in Table 7.2.1.3.1-1.

The inherent CCD memory configuration is the shift register with serial-in/serial-out operation. There are currently three basic types of CCD memory organization: the serpentine configuration, the series-parallel-series (SPS) configuration, and the line-addressable random-access memory (LARAM) configuration. Some other organizations derived from these basic configurations include the multiplexed electrode per bit, the interlaced SPS, and the addressed drum type structure.

The serpentine organization is obtained by connecting several sequential multi-stage shift registers in series by means of refresh cells. Internally, all data bits are sequentially shifted through all cells of the memory. An advantage of the serpentine configuration is that it permits construction of very long registers with excellent low frequency operational characteristics. A disadvantage is that the power required is higher at higher shift rates than for the other approaches because all bits are moving at the same frequency.

The series-parallel-series (SPS) organization provides the advantages of greater packing density of the basic CCD shift register with lower power consumption. The input data in the SPS organization are first shifted into a horizontal register. After the horizontal register is filled with data, the data are transferred in parallel to several vertical registers. Thus the shift rate of the vertical registers is N^{-1} that of the horizontal input register, where N is the number of bits in the horizontal register. In the SPS organization, a separate horizontal output register is normally used. An advantage of this configuration is low power dissipation that results from the slow vertical clock rate. The main disadvantage of this configuration is the long delay between input and output. This limits the low-frequency and high-temperature

TABLE 7.2.1.3.1-1. STATE OF THE ART IN CCD MEMORY

CHARACTERISTIC	1977	1975
Chip		
Capacity	65,536 bits	16,384 bits
Size	40K mil^2	44K mil^2
Organization	16, 4-Kbit SPS Loops*	LARAM
Maximum Latency	820 μsec (at 5 MHz)	32 μsec (at 16 MHz)
Maximum Data Transfer Rate	5 Mbits/sec	16 Mbits/sec
Power - Read/Write (at 5 MHz)	260 mW (at 5 MHz)	200 mW (at 16 MHz)
- Standby	25 mW	50 mW
Packaging	16 pin dip	22 pin dip
Cost	0.1 ¢/bit	0.2 ¢/bit
MTBF	NA	NA
System		
Capacity	4 Mbytes/chassis	2 Mbytes/chassis
Volume	NA	NA
Maximum Access Time	1 msec	40 μsec
Transfer Rate**	5 Mbytes/sec	16 Mbytes/sec
Cost	0.15 ¢/bit	0.4 ¢/bit
Error Rate	NA	NA
MTBF	NA	NA

*Each addressable 4,096-bit loop organized as 32 by 127 SPS.
**System design dependent.

operation of the memory because of leakage currents. This configuration does, however, offer extremely fast data rates and is therefore an attractive configuration for the replacement of mechanical magnetic storage devices.

The line-addressable CCD configuration combines an MOS address selection matrix with a number of sequential CCD shift registers. Only the data in the selected line are shifted at a time to provide input, output, or refresh. The line-addressable configuration combines fast access times with low power consumption.

7.2.1.3.2 Trends in CCD Memory - CCD memory devices have only been commercially available since 1974; however, the technology is already rapidly moving toward increased capacity and lower cost per bit. The major application for CCD memory devices in the future will be in the production of "electronic disk" systems to replace and supplement mechanical mass storage devices such as fixed-head disks. Currently, it is not clear whether CCD or magnetic bubbles will eventually dominate this market. CCD devices currently provide approximately a 10:1 speed advantage compared with bubble devices. However, CCD devices require power during standby whereas magnetic bubbles are nonvolatile.

Figure 7.2.1.3.2-1 illustrates chip capacity trends for CCD memory devices. The projected trend indicates a continuation of the yearly doubling of semiconductor memory chip capacity that began in 1968. The application of electron-beam lithographic and direct wafer fabrication techniques should enable the current trend to continue through 1985.

Electron-beam hardware for generating masks with 1-μm resolution routinely are currently being built and are expected to be in use prior to mid-1978. Direct production of 4-in. wafers with resolution to approximately 0.5 μm is technically feasible for special-purpose applications, but is not economically feasible for production applications. Direct wafer fabrication is not likely to be feasible prior to the next 4 or 5 years. In the meantime, improvements in optical lithography and hybrid electron beam/lithographic systems will provide the resolutions required to advance at the rates projected in Figure 7.2.1.3.2-1.

Although CCD cost trends are not clearly established based on historical data, the cost of future CCD devices will closely parallel chip capacity and density trends. Figure 7.2.1.3.2-2 presents CCD memory cost trends for CCD chips and memory systems.

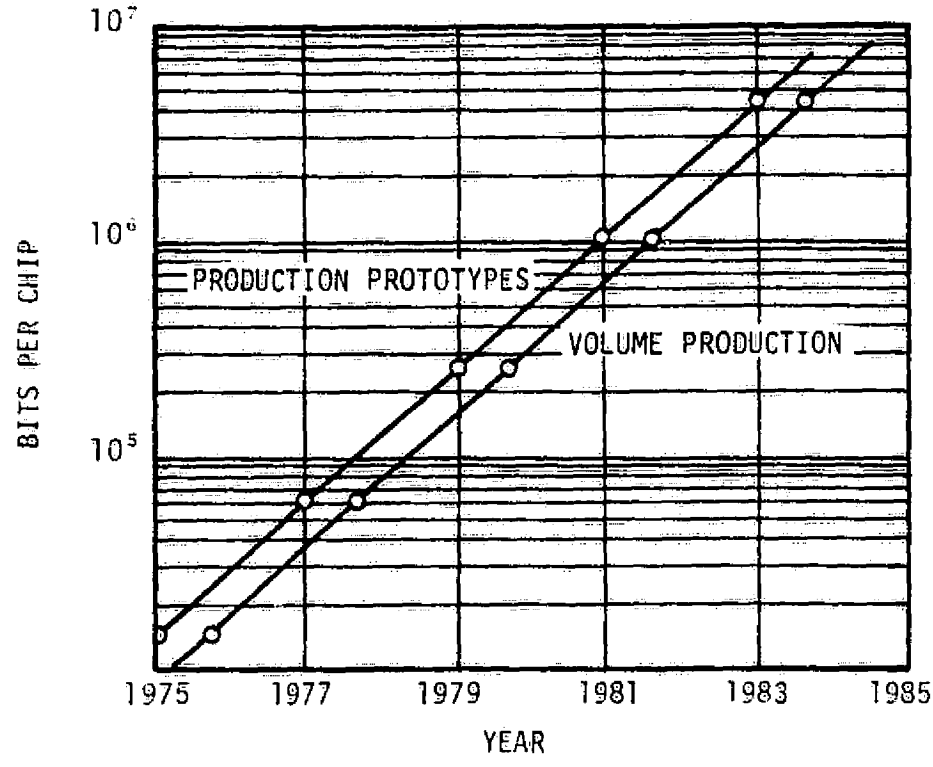


FIGURE 7.2.1.3.2-1. TRENDS IN CCD MEMORY CHIP CAPACITY

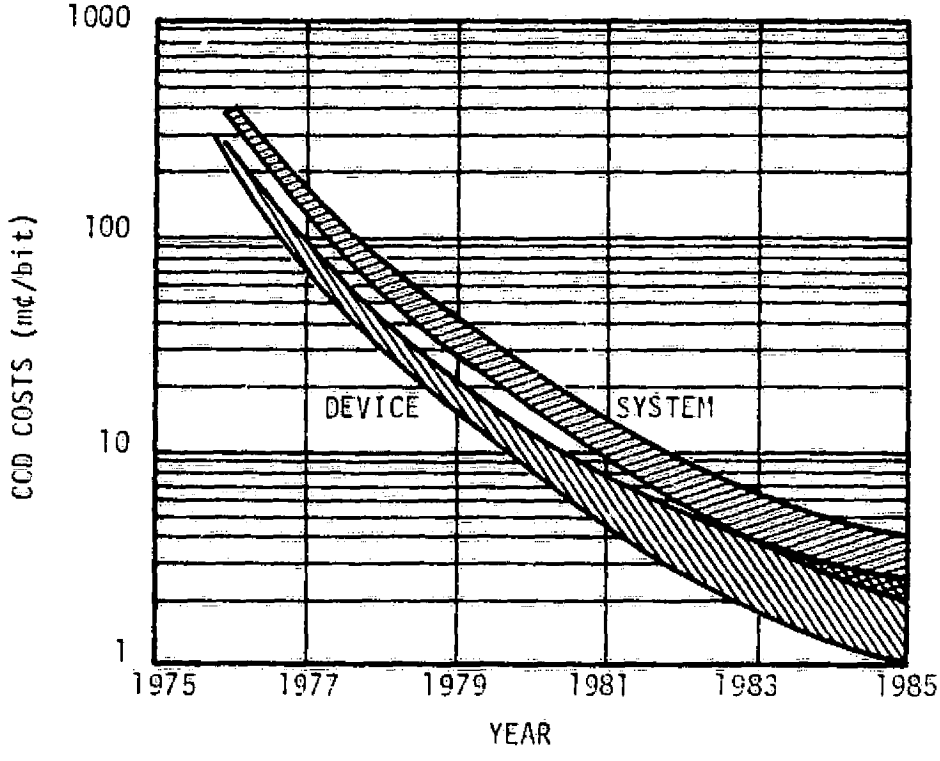


FIGURE 7.2.1.3.2-2. TRENDS IN CCD MEMORY COST

7.2.1.3.3 Projected Developments in CCD Memory - Some authorities predict that by 1980 there will be more CCD memory capacity shipped per year than any other semiconductor technology at a cost of less than 10 millicents per bit. Projected characteristics for CCD memory devices that will be commercially available in 1980 and 1985 are presented in Table 7.2.1.3.3-1. Because CCD memory devices are inherently serial, several significantly different architectures will evolve with significant differences between the performance and cost of the devices that provide the highest density and the devices that provide the fastest access.

In addition to the conventional shift register approach to CCD memory design, charge coupling has been used in combination with MOS circuitry to produce novel memory and logic element designs. Texas Instruments recently disclosed an experimental CCD RAM cell device (Refs. 7-14 and 7-15) that stores data in switchable CCD capacitor regions implanted beneath an MOS gate. Thus the entire memory cell is no larger than a single gate that can measure less than half a mil square. The CCD RAM saves even more chip space by requiring only two access lines per cell, whereas all other currently produced RAMs require three lines per cell. Since each CCD RAM cell only has two access lines, each cell is a component that is less complex than a transistor. Therefore, the CCD RAM is also called a no-transistor RAM.

IBM has recently reported a similar experimental device (Ref. 7-16) called a merged charge memory (MCM), with a cell size of 0.3 mil square. These laboratory developments indicate that CCD RAM type memory cells may be used to fabricate next-generation (65,536-bit) NMOS RAM devices.

TABLE 7.2.1.3.3-1. PROJECTED DEVELOPMENTS IN CCD MEMORY

CHARACTERISTIC	1980	1985
Chip		
Capacity	256 Kbits*	4 Mbits*
Size	40K mil^2	50K mil^2
Organization	NA	NA
Maximum Latency	1 msec*	1 msec*
Operating Frequency	1 to 10 MHz	1 to 20 MHz
Power - Read/Write	300 mW	500 mW
- Low	50 mW	100 mW
Packaging	16 pin dip	TBD
Cost	0.01 ¢/bit*	0.0015 ¢/bit*
MTBF	NA	NA
System		
Capacity	8 Mbytes/chassis	128 Mbytes/chassis
Volume	NA	NA
Maximum Access Time	1 msec	1 msec
Transfer Rate	Up to 80 Mbits/sec	Up to 160 Mbits/sec
Cost	0.02 ¢/bit	0.003 ¢/bit
Error Rate	NA	NA
MTBF	NA	NA

*Lower capacity, higher speed devices will be available at higher cost/bit.

7.2.2 Other Electronic/Optical Memory

7.2.2.1 Electron Beam Addressed Memory -

7.2.2.1.1 State of the Art in Electron Beam Addressed Memory -

The use of electron beams for accessing memory has been under investigation for a number of years, principally by Stanford Research Institute, General Electric, and Micro-Bit Corporation. Currently, only Micro-Bit is still active in the field. Such memories are called either Electron Beam Addressable Memories (EBAM) or Beam Addressable MOS Memories (BEAMOS). The technology is based on the use of an electron beam to read and write data on a simple, unstructured MOS chip. Addressing the two-dimensional array of data is accomplished by deflection of the electron beam. Data are stored on the MOS target as the presence or absence of positive charge in the oxide near the oxide-silicon interface. Charge storage is accomplished by the application of a bias voltage to the target while the beam is swept along a track during a write or erase cycle. In a subsequent read cycle, the beam is swept over the same track, and where a charge is present a current is produced in a sense circuit. If no charge is present, no significant current is produced.

A complete BEAMOS or EBAM memory system consists of one or more tubes, address and interface logic, control circuits, and power supplies. The tube contains the MOS target and electron beam addressing system enclosed in a sealed evacuated envelope. Tubes are being designed with both one and two lens-deflector fields. The number of bits that can be accessed in a one-stage tube is limited by aberrations in the deflector and by inaccuracies and instabilities of the deflection eletronics. These limitations can be pushed outward by several orders of magnitude through the use of two-stage deflection and an array of lenses known as the fly's eye configuration.

Development of BEAMOS or EBAM technology has progressed to the prototype stage, with units scheduled to be commercially available in 1980. Table 7.2.2.1.1-1 summarizes the characteristics of the prototype units and memory systems. A memory access that invokes an operation change (e.g., read to write) requires a change in the bias of the oxide. Therefore, as indicated in the table, this type of access takes more time than

TABLE 7.2.2.1.1-1. STATE OF THE ART IN ELECTRON BEAM ADDRESSED MEMORY

CHARACTERISTICS*	ONE-STAGE	TWO-STAGE
Tube		
Dimensions (length by diameter)	26 by 5.5 cm	42 by 10 cm
Capacity	4 Mbits	32 Mbits
Storage Density	8.3×10^5 bits/cm^2	4.2×10^6 bits/cm^2
Cost	0.005 ¢/bit	0.005 ¢/bit
System		
Capacity	Up to 64×10^6 bits	Up to 6×10^8 bits
Access Time - No Operation Change	5 μsec	5 μsec
Access Time - Operation Change	10 to 20 μsec	20 to 30 μsec
Service Time	440 μsec	440 μsec
Recording Data Rate**	0.5 to 5 Mpbs	5 to 10 Mbps
Power	250 W	250 W
Temperature Range	-50 to +70°C	-50 to +70°C
Radiation Tolerance	10^5 rad	10^5 rad
MTBF	8,000 hr	8,000 hr
Cost+	0.1 ¢/bit	0.1 ¢/bit

*Laboratory models; first systems to be delivered early 1980.

**Higher read rate possible; higher write rate if tubes are operated in parallel.

+Upon entry into market; will fall to 0.02 to 0.05 ¢/bit as volume increases; cost listed is for large, multitube system.

successive calls of the same operation, which are limited primarily by the switching and settling time of the deflection amplifier. The service time listed in the table is the average time between the acceptance by the memory of a new command and the earliest time that the next command can be accepted. The recording data rate for a tube is determined by the beam current available. Because considerably less beam current is required for a read operation, reading can be performed at a higher rate than writing.

Electron beam memories require very stable and highly regulated power supplies and a very fast and accurate deflection electronics system, both of which contribute significantly to the overall system cost. For this reason it is highly desirable to build multitube systems in which a large portion of this electronics is shared. Only by sharing the electronics is the cost per bit reduced to acceptable levels.

The nonstructured nature of the storage plane and the flexibility of electron beam addressing allow the MOS target to be formatted in a wide variety of ways. BEAMOS or EBAM systems can be used to access a single bit at a time by allowing the deflection system to settle at a particular location and then turning on the beam to interrogate the charge stored at that target position. However, it is highly probable that most systems will be used in a block-oriented mode in which information is written as an encoded string of transitions along a continuous path. Random access to the start of the block is achieved digitally, and the block itself is scanned on the fly with the beam on continuously. This is a more efficient organization because access time is primarily dependent on the deflection settling time, which in the block organization is incurred only for the starting location of a block. The block length can vary over wide limits and there are no built-in restrictions on word length. Therefore, redundancy can be included easily within each block for error control to enhance system reliability.

The MOS target is a relatively nonvolatile storage medium. It stores data for several weeks, even when power is off, but it does not store data permanently because of signal decay. Signal decay is very slow with the power off or with the power on and zero or negative oxide bias

values. With the power off, the measured signal is 90% of its initial value after five days of storage and about 80% after one month (Ref. 7-17). Decay is somewhat faster at positive bias. In addition, the signal is reduced slightly by each readout. As a result of these sources of decay, refreshing of the memory is required, even when the memory is not being used. When reading at 10 Mbits/sec, a rewrite becomes necessary after 10 to 20 reads. At higher read data rates, the number of reads before rewrite decreases.

Because the electron beam is a source of radiation, it causes a gradual permanent degradation of the target. This degradation limits the lifetime in each location to the order of 10^7 write-erase cycles. To lessen the effects of this target fatigue, present BEAMOS or EBAM systems rotate or permute the data to average the usage of the target area. Only if the tube itself is replaced on a preventive maintenance schedule can the MTBF given in Table 7.2.2.1.1-1 be achieved.

BEAMOS or EBAM memories exhibit many characteristics required for military environment applications. A major advantage is the fact that the systems are all-electronic and completely sealed. No parts are especially sensitive to vibration and the section most important for beam positioning, the matrix lens/target assembly, is very rugged. Tests of the storage plane over the -50°C to +70°C temperature range indicate that charge storage is not greatly affected by temperature. In this temperature range there should be less than 0.1-μm variation in the position of the addressing beam. Although the memory requires a shield to protect against external fields, a shield weighing only 10 lb should protect against fields up to 10 g in strength. The memory plane is quite tolerant of ionizing radiation. Ionizing radiation of 10^5 rads causes no more than 10% reduction in signal level, while no permanent radiation damage has been observed up to 6×10^7 rads.

7.2.2.1.2 Trends in Electron Beam Addressed Memory - While functional BEAMOS or EBAM memory systems exist today, they are far from "home free". To lower costs while increasing capacity and performance, two significant problem areas must be addressed. First, electron optics must be upgraded to accomplish submicron beam diameters in production so that density goals can be realized. Second, cathode development must occur to increase beam brightness as the beam is made smaller to achieve greater density. Assuming that progress is made in these areas, the trends stated in the following paragraphs should be realized.

BEAMOS or EBAM memories offer great potential in terms of performance, cost per bit, and total capacity. System performance is expected to approach that of main memory with a capacity approaching and even exceeding that of the large moving-head disk files. Therefore, these memory systems have the potential not only of bridging the memory access gap but also of eliminating it and thus effecting a dramatic reduction in operating system software complexity and cost.

During the next decade BEAMOS or EBAM memories will benefit from research in both semiconductor and electron beam technologies. As the result of improvements in MOS target parameters, the use of two-stage deflection in the array optics configuration, and increases in the addressable target area of a single tube arising from advances in electron optics, storage densities will increase and access times will decrease, as illustrated in Figures 7.2.2.1.2-1 and 7.2.2.1.2-2, respectively. At the same time, recording data rates will increase, as illustrated in Figure 7.2.2.1.2-3. The increases in storage density and addressable target area will bring increases in tube capacity, as illustrated in Figure 7.2.2.1.2-4. As a result of these technology improvements, cost per bit will decline rapidly, as illustrated in Figure 7.2.2.1.2-5. In addition, signal decay and target degradation improvements will be realized and work on one-stage deflection will decrease except for use as a study tool.

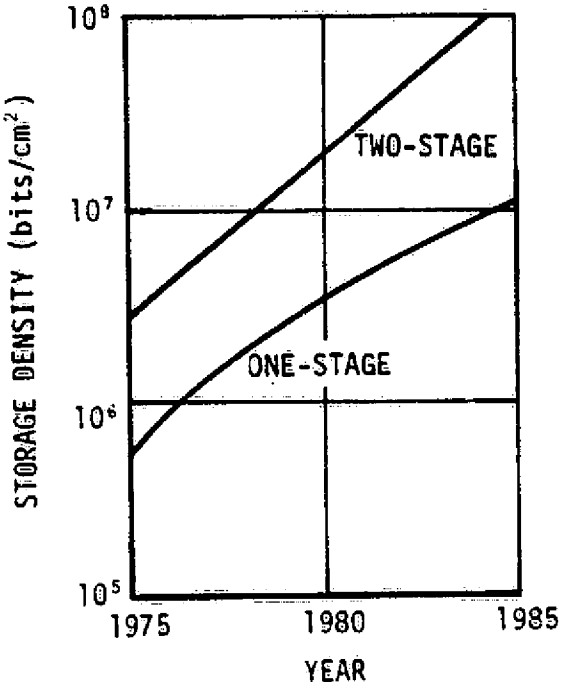


FIGURE 7.2.2.1.2-1. ELECTRON BEAM ADDRESSED MEMORY STORAGE DENSITY TRENDS

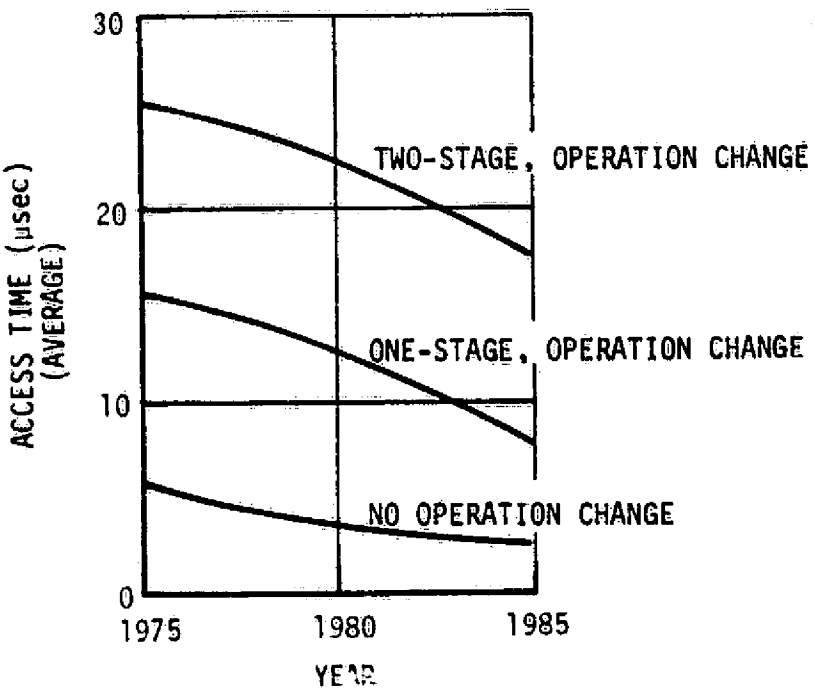


FIGURE 7.2.2.1.2-2. ELECTRON BEAM ADDRESSED MEMORY ACCESS TIME TRENDS

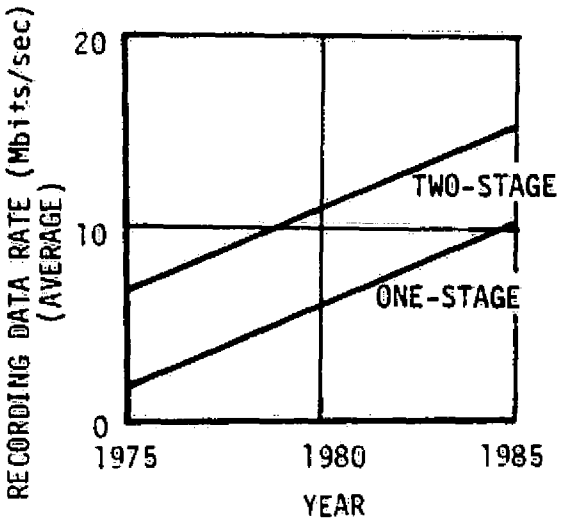


FIGURE 7.2.2.1.2-3 ELECTRON BEAM ADDRESSED MEMORY RECORDING DATA RATE TRENDS

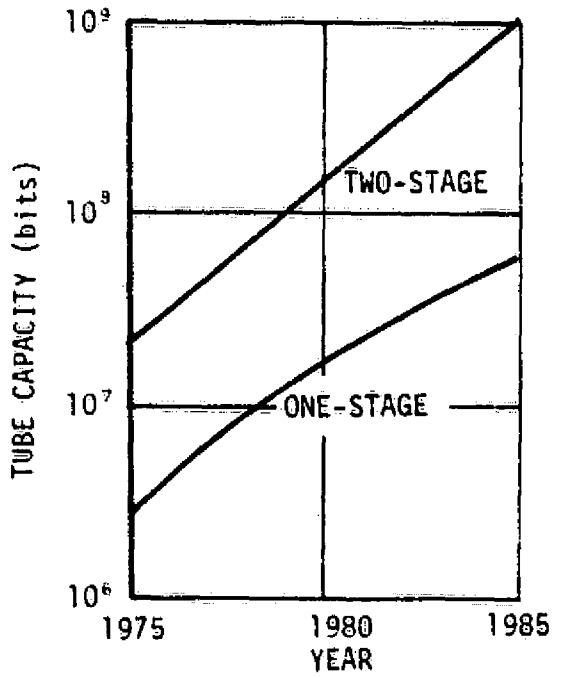


FIGURE 7.2.2.1.2-4 ELECTRON BEAM ADDRESSED MEMORY TUBE CAPACITY TRENDS

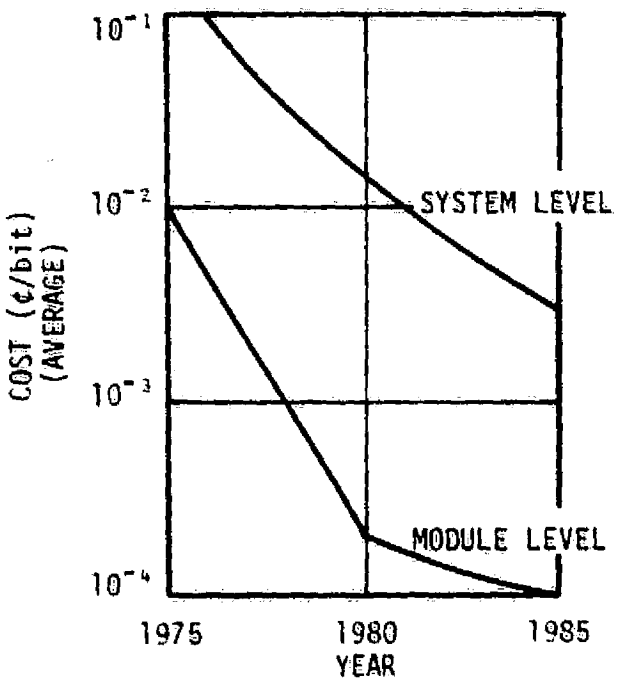


FIGURE 7.2.2.1.2-5. ELECTRON BEAM ADDRESSED MEMORY COST TRENDS

7.2.2.1.3 Projected Developments in Electron Beam Addressed Memory - Assuming that progress is made in the problem areas discussed under trends (Subsection 7.2.2.1.2), BEAMOS or EBAM memory systems should be commercially available in 1980 and 1985 with the characteristics presented in Tables 7.2.2.1.3-1 and 7.2.2.1.3-2, respectively. The minimum system capacity is equal to the tube capacity. Since system-level cost is dependent on the amount of sharing of electronics that is achieved, multitube systems will be dominant. The system costs given in the tables are for these multitube systems. Emphasis will be placed on utilizing two-stage deflection (to the practical exclusion of one-stage deflection) and an array of lenses known as the fly's eye configuration because of the improvement it offers in terms of addressable target area. The ultimate theoretical limit on packing density is approximately 10^{10} bits/cm^2 (Ref. 7-17). The development of tube capacities much greater than 10^9 bits will require the use of field-emitter cathodes that have three to five orders of magnitude greater brightness than the currently used thermal dispenser cathode. Such cathodes have been developed already as a result of the demand for higher-resolution scanning electron microscopy (Ref. 7-18).

The initial impact of the electron beam memory technology will be as a replacement for the fast-access auxiliary storage devices, currently dominated by fixed-head disks. The BEAMOS or EBAM systems will probably be cost-competitive with the high-performance fixed-head disks while offering significant performance improvement. Eventually these systems will be cost-competitive with all on-line random access peripheral memories but with far superior performance. Equally important is their potential use as main memory extensions, in combination with semiconductor cache, where they will have a large price advantage at comparable performance. Amdahl is on record as being willing to replace virtual disk memory, with its big loss of throughput, with electron beam memories, or at least electron beam cache, if they prove cost-effective.

7.2.2.2 Magnetic Bubble Memory - Magnetic Bubble Memory is discussed in Section 3.2.

TABLE 7.2.2.1.3-1. PROJECTED TECHNOLOGY IN ELECTRON BEAM ADDRESSED MEMORY, 1980

CHARACTERISTICS	ONE-STAGE	TWO-STAGE
Tube		
Dimensions (length by diameter)	26 by 5.5 cm	42 by 10 cm
Capacity	16 Mbits	128 Mbits
Storage Density	3.3×10^6 bits/cm^2	1.7×10^7 bits/cm^2
Cost	0.0002 ¢/bit	0.0002 ¢/bit
System		
Capacity	Up to 256 Mbits	Up to 2 Gbits
Access Time - No Operation Change	3.3 μsec	3.3 μsec
Access Time - Operation Change	10 to 15 μsec	20 to 25 μsec
Service Time	NA	NA
Recording Data Rate*	4 to 8 Mbps	10 to 12 Mbps
Power	NA	NA
Temperature Range	-50 to +70°C	-50 to +70°C
Radiation Tolerance	10^5 rads	10^5 rads
MTBF	NA	NA
Cost**	0.01 to 0.02 ¢/bit	0.01 to 0.02 ¢/bit

*Higher read rate possible; higher write rate if tubes are operated in parallel.

**For large, multitube system.

TABLE 7.2.2.1.3-2. PROJECTED TECHNOLOGY IN ELECTRON BEAM ADDRESSED MEMORY, 1985

CHARACTERISTICS	ONE-STAGE	TWO-STAGE
Tube		
Dimensions (length by diameter)	26 by 5.5 cm	42 by 10 cm
Capacity	53 Mbits	1 Gbit
Storage Density	1×10^7 bits/cm^2	1.3×10^8 bits/cm^2
Cost	0.0001 ¢/bit	0.0001 ¢/bit
System		
Capacity	Up to 848 Mbits	Up to 8 Gbits
Access Time - No Operation Change	2.5 μsec	2.5 μsec
Access Time - Operation Change	5 to 10 μsec	15 to 20 μsec
Service Time	NA	NA
Recording Data Rate	10 Mbps	15 Mbps
Power	NA	NA
Temperature	-50 to +70°C	-50 to +70°C
Radiation Tolerance	10^5 rads	10^5 rads
MTBF	NA	NA
Cost**	0.001 to 0.005 ¢/bit	0.001 to 0.005 ¢/bit

*Higher read rate possible; higher write rate if tubes are operated in parallel.

**For large, multitube system.

REPRODUCIBILITY OF THE ORIGINAL PAGE

7.2.3 Core Memories

Computer mainframe memories used magnetic core technology almost exclusively from the early 1950's up until IBM introduced semiconductor memory in the early 1970's. During this time the demise of core memory was repeatedly predicted. In the late 1950's it was thought that cryogenics would supplant core; in the early 1960's magnetic thin film; in the later 1960's plated wire; and finally semiconductors.

Although semiconductor memory has replaced core as the mainframe memory for large computer systems, minicomputer manufacturers continue to ship large quantities of core memories for process control applications where the nonvolatility of core is considered important. Another current application of core memory is as a bulk storage medium to fill the "access gap" between the speed of mainframe semiconductor memory and that of fixed-head disk and other peripheral storage devices.

Core is the principal storage medium used for airborne and spaceborne data systems; however, plated wire is also used extensively where low power operation is required. Plated wire memory, unlike core, provides nondestructive readout (NDRO); thus the data does not have to be rewritten each time the memory is read. This results in faster cycle times and lower power consumption. Another alleged advantage of NDRO memories is that the data is less volatile because it cannot be altered by noise during the rewrite process.

Some types of core memory designs such as read-only core memories and NDRO core memories are no longer used. Read-only core memories were used for microprogrammable CPU control prior to the development of semiconductor read-only memories. The read-only core memory used the presence or absence of a core in a linear-select, two-dimensional array to store data. NDRO core memories utilized cores with two holes instead of one. One of the major applications for NDRO memories is in the design of format controllers for airborne and spaceborne data acquisition systems. During the 1960's the NDRO memory market was captured by plated wire. More recently, a large percentage of the fixed program memory market has been taken by erasable and nonerasable, programmable, semiconductor read-only memories.

7.2.3.1 State of the Art in Core Memories - Core memory systems are currently used for three types of data storage applications:

- Microcomputer/Minicomputer Memory
- Core-Based Mass Storage
- Airborne Computer Memory.

Table 7.2.3.1-1 presents typical characteristics for each of these types of core memory. Because many of the current systems package a complete memory system as a single module, there may be some confusion as to the difference between a module and a system. A system is defined as one or more modules packaged as a single unit.

State-of-the-art core memory system designs utilize either two-and-one-half-dimensional or three-dimensional organization. In the three-dimensional systems the cores are arranged in an X-Y matrix called a bit plane. M of these bit planes are required to implement a full memory. The bit planes are interconnected by wiring the X and Y wires of each plane in series, creating a three-dimensional array called a core stack. The basic three-dimensional memory requires $(N/M)^{\frac{1}{2}}$ "X" drivers and $(N/M)^{\frac{1}{2}}$ "Y" drivers plus M inhibit drivers, where N is the total number of bits in the memory. The number of drivers can be further reduced by arranging the X and Y driver each in a matrix, thus reducing the total number of drivers from $2(N/M)^{\frac{1}{2}}$ to $4(N/M)^{\frac{1}{4}}$ plus an additional $4(N/M)^{\frac{1}{2}}$ diodes. Current three-dimensional technology uses the same line for sense and inhibit, thereby reducing the number of wires per core from 4 to 3.

The two-and-one-half-dimensional organization is similar to the three-dimensional organization in that the X drive lines are wired in series. The two-and-one-half-dimensional organization differs from the three-dimensional organization in that separate sets of Y drivers are used for each plane. In the conventional two-and-one-half-dimensional organization, the readout signal is sensed on a third sense wire that links all cores in a bit plane. The two-and-one-half-dimensional organization has been used to make very low-cost, core-based, mass storage (also called bulk core). In this type of memory, the separate sense wire is eliminated and the readout signal is sensed on M sets of Y drive

TABLE 7.2.3.1-1. STATE OF THE ART IN CORE MEMORIES

CHARACTERISTICS	MICROCOMPUTER/ MINICOMPUTER MEMORY	CORE-BASED MASS STORAGE	AIRBORNE COMPUTER MEMORY
MODULE			
Capacity (bytes, minimum to maximum)	1K to 64K	256K to 4,096K bytes	16K to 64K
Access Time (nsec)	275	600	350
Cycle Time (nsec)	650	1,500	900
Power Dissipation (μW/bit)	N/A	40	250
Configurations	3D (3-Wire)	2½D (2-wire) and 3D (3-Wire)	2½D (2-Wire) and 3D (3-Wire)
Physical Size (in.)	16 × 11.5 × 1.0	N/A	1.5 × 6.0 × 9.0
Weight (lb)	4	N/A	3.2 (less power supply)
MTBF (hr)	Up to 80,000	N/A	20,000
SYSTEM			
Capacity Bytes (minimum to maximum)	1K to 1,024K	256K to 4,096K	16K to 256K
Cost (¢/bit)	0.3	0.2	2.0
Access Time (nsec)	275*	2,000	350*
Cycle Time (nsec)	650*	4,000	900*
Power Dissipation (μW/bit)	210 (full capacity)	22 (full capacity)	**
Physical Size (in. maximum)	23 × 19 × 18	23 × 19 × 22	**
Maximum Weight (lb)	160	150	**
MTBF	**	**	**
Environment - Operating Nonoperating	0 to 55°C -55 to +85°C	0 to +55°C -55 to +80°C	-55 to 85°C N/A

*Measured at memory module interface connector
**Depends on system capacity
N/A - Not available

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

wires. The increase in the complexity of the sensing electronics is more than offset by the cost saving realized by eliminating the continuous long sense wire. The Ampex Megastore, core-based, mass storage system utilizes the two-and-one-half-dimensional, two-wire technology.

Some of the factors responsible for improvements in core memory cost/performance during recent years include advances in ferrite technology, more efficient stack configurations, and greater use of integrated circuitry. Ferrite technology improvements include the use of smaller cores, the discovery of temperature-independent materials, faster materials, and simplified core manufacturing techniques. Currently, cores as small as 12 or 13 mils in outer diameter and 7 mils inner diameter are being wired. Except for IBM, which uses automatic core wiring equipment for its 4 Pi computer core memories, manufacturers of core memories use manual labor in overseas facilities for wiring core planes. One of the recent innovations in core manufacturing technology is the use of a tape punch process, rather than the powder compaction process, to economically produce high-quality, 13-mil, ceramic, temperature-independent cores. State-of-the-art, temperature-independent cores provide stable operation over a temperature range of -55 to +125°C without the necessity of compensating the drive currents as was required in earlier core system designs, thus simplifying the electronic drive circuitry.

Improvements in core stack designs have resulted in increased stack density (currently approximately 6,500 cores/in^2) and more cores per sense line (currently up to 32,768). Increased stack density has made possible the packaging of core memory systems in the low-cost planar configurations rather than the more costly "stacked" stack configurations that were previously used.

Core memory electronics costs have decreased in recent years as a result of the use of MSI and LSI circuitry, greater number of cores per sense winding, and the temperature-independent core materials.

7.2.3.2 Trends in Core Memories - The current trends in core memory development are directed toward achieving improved performance and lower cost per bit of storage. Figure 7.2.3.2-1 shows how the cost of commercial core memory has dropped from approximately 10¢/bit in 1960 to currently less than 0.3¢/bit. This trend should continue through 1985 if no suitable core replacement such as the NDRO semiconductor is developed.

Figure 7.2.3.2-2 illustrates how core memory system speeds have improved since 1960. Further improvements in both access times and cycle times can be expected as smaller cores (8-mil) are developed that require lower drive currents.

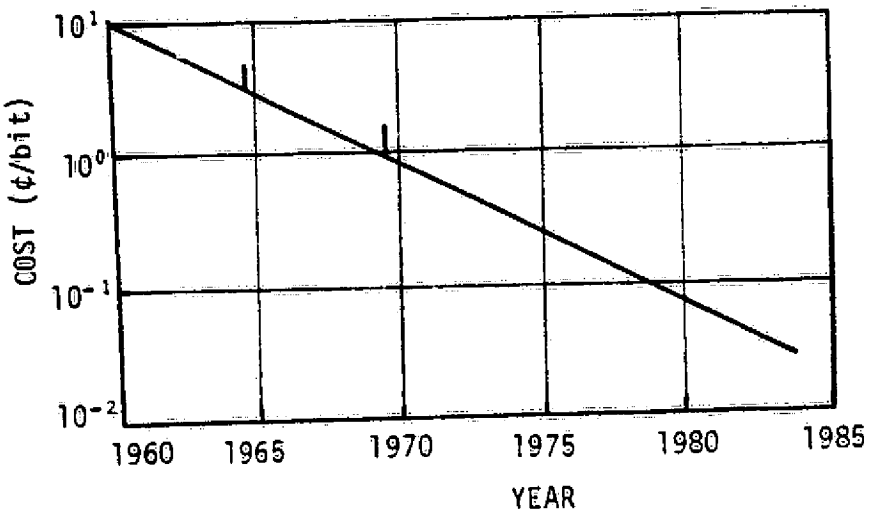


FIGURE 7.2.3.2-1. TRENDS IN CORE MEMORY SYSTEMS COST

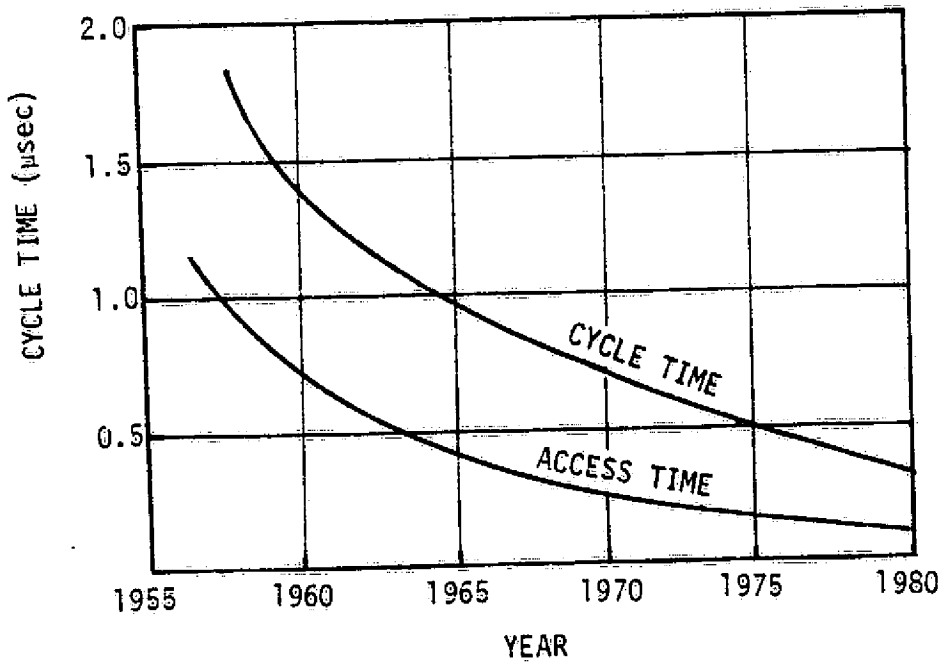


FIGURE 7.2.3.2-2. TRENDS IN CORE MEMORY SYSTEM SPEED

7.2.3.3 Projected Developments in Core Memory - The trends in core memory technology indicate that significant improvements in both cost and performance can be expected by 1985, provided a replacement technology such as NDRO semiconductor memory does not become commercially available at lower cost. Table 7.2.3.3-1 presents projected characteristics of commercial core memories for 1980 and 1985. As seen from Table 7.2.3.3-1, the cost per bit for core storage is expected to continue dropping through 1985 as a result of eliminating labor-intensive manufacturing procedures. There will also be improvements in speed, reliability, and storage density. Improvements in speed and storage density will result primarily from the use of smaller cores. Reliability improvements will result from the use of more LSI and MSI circuitry.

Another projected core memory development is a modular core memory device that Ampex calls the "core chip". In this device the cores are lodged in a ceramic "cavity plate" that also serves as a substrate for deposited wiring that replaces the threaded wires normally used in thre-dimensional organized core memories. Ampex is expected to announce a production 1K device later this year.

TABLE 7.2.3.3-1. PROJECTED TECHNOLOGY IN COMMERCIAL CORE MEMORIES

CHARACTERISTICS	1980	1985
Minicomputer Module		
Capacity (bytes, minimum to maximum)	1K to 128K	1K to 256K
Access Time (nsec)	200	100
Cycle Time (nsec)	500	300
Power Dissipation (mW/bit)	150	100
Configurations	2½D (2-Wire) and 3D (3-Wire)	2½D (2-Wire) and 3D (3-Wire)
Stack Density (bits/in^2)	6,500	12,000
Core Size	12 mils	8 mils
MTBF (hr)	10^5	2×10^5
Cost (¢/bit)	0.10	0.035
Mass Storage System		
Capacity	256K to 4,096K	256 to 8,192K
Access Time (μsec)	1.5	1.25
Cycle Time (μsec)	3.0	2.5
Maximum Physical Size (in.)	23 × 19 × 22	23 × 19 × 22
Cost (¢/bit)	0.10	0.035

7.2.4 Disks

7.2.4.1 Fixed-Head Disks -

7.2.4.1.1 State of the Art in Fixed-Head Disks - Fixed-head disks (FHDs), also called head-per-track disks, are configured such that stationary, rigidly mounted read/write heads are positioned over the rotating magnetic surface and a head reads or writes only on one track. Table 7.2.4.1.1-1 contains the state-of-the-art system characteristics for fixed-head disks.

Industry currently pays at least a 30:1 price penalty between fixed-head and moving-head disks to get less than a 10:1 performance improvement. Consequently, FHDs are generally used only in applications where the slower access time of movable head disks cannot be tolerated. The FHD average access time (latency time) to a given block address is one-half the rotational period of the disk and is typically about 8.3 msec.

The magnetic medium currently used in disk technology is a magnetic oxide particulate coating about 1 μm thick. The read/write heads are made of a ceramic/ferrite material with a gap of 1 to 2 μm and spaced 1 μm or less from the disk surface. As linear and track densities increase, the thickness of the magnetic coating, head-to-surface gap, mechanical tolerances of the rotational spindle, and the head spacing will have to be decreased.

During the last 2 years, the increased demand by users for greater storage and faster access has increased the use of FHDs because advances in areal density have produced rather large capacities at reduced cost. With the increased desire for large-capacity systems, the concept of interchangeability is no longer a necessary or economical feature. The advantages of fixed-head disks are:

- No expense of disk-loading hardware or mechanism to maintain tolerance and contamination protection because the disk surfaces and heads are hermetically sealed in one container
- No data package limitations imposed by the necessity for maintaining the radial alignment tolerance necessary for drive-pack interchangeability

TABLE 7.2.4.1.1-2. STATE OF THE ART IN FIXED HEAD DISKS

SYSTEM CHARACTERISTICS	FIXED-HEAD DISKS
Track Density	300 to 500 tracks/in
Linear Density	5,636 to 7,000 bits/in
Media Areal Density	Up to 3.5 Mbits/in^2
Drive Capacity	0.5 to 472.5 Mbytes
Maximum System Capacity	Up to 647.5 Mbytes
Physical Size	19 by 23 by 17 to 42 by 33.5 by 46.5 in.
Weight	70 to 900 lb
Operating Temperature Range	0 to 50°C
Humidity Tolerance	10 to 90%
Rotation Time	5 to 10 msec
Average Access Time	5 to 13 msec
Soft Error Rate	10^{-11}
Hard Error Rate	10^{-13}
Drive Cost	0.021 to 0.90 ¢/bit
System Cost	0.06 to 1.0 ¢/bit
MTBF	10,000 hr

- No expense for accurate angular alignment of read/write heads to prevent problems due to gap skew
- High reliability of mechanical portions of disk drive
- High data transfer rates and fast access time.

7.2.4.1.2 Trends in Fixed-Head Disks - A recent design trend in fixed-head storage combines fixed-head files on the same spindle with moving-head files and thereby reduces the cost premium of fixed-head storage and improves the access time for moving-head storage. Although this innovation has revived the use of FHD technology, the cost is still sufficient to be challenged in the near future by solid-state storage technology.

One area that will show little significant improvement for FHD is access time. A large improvement does not appear feasible or worthwhile since rotational speed plays a large part in determining access time, and significant increases in rotational speeds will not be cost-effective. Advanced error correction codes, such as cyclic Bose-Chaudhuri-Hocquenghem (BCH) codes, which have excellent burst error correction and detection capabilities, are expected to be in use by 1980.

The areal density trends of FHDs are given in Figure 7.2.4.1.2-1. This figure shows that a significant increase in areal density is possible in the next decade. Figure 7.2.4.1.2-2 gives the FHD system cost trends. The cost per bit of storage for FHDs will not decrease as much as moving-head disks because the cost of the disk assembly and heads, including the controller, will remain fairly constant for the next 10 years. The current price trends of FHD have shown only a decrease of 10% to 20% per year, as compared with MHD's 20 to 25%. Even though disk density trends will increase dramatically, the end is in sight and the technological limit for densities in magnetic media will probably be reached this century.

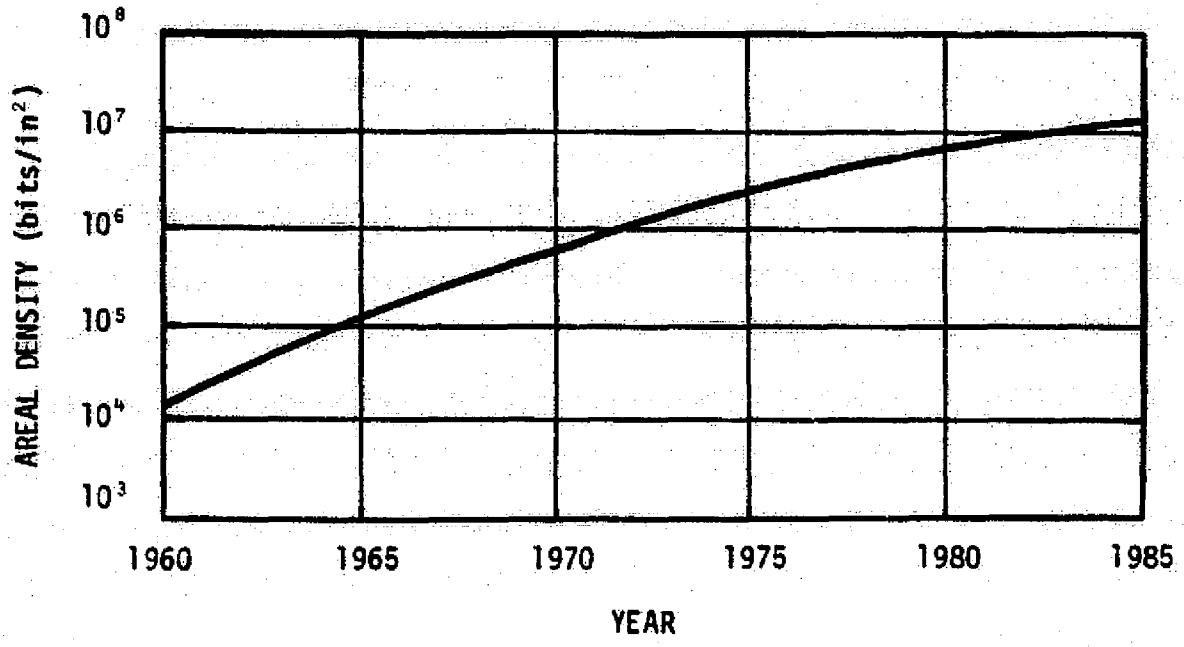


FIGURE 7.2.4.1.2-1. TRENDS IN FIXED-HEAD DISK AREAL DENSITY

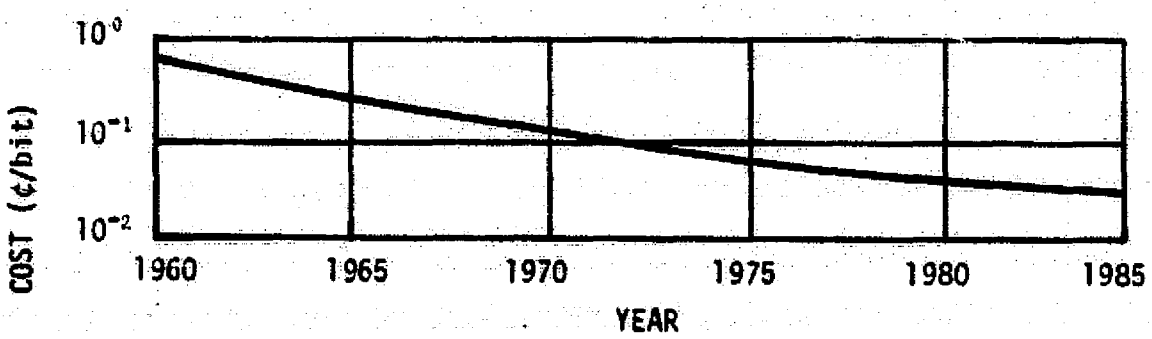


FIGURE 7.2.4.1.2-2. TRENDS IN FIXED-HEAD DISK SYSTEM COST

7.2.4.1.3 Projected Developments in Fixed-Head Disks - During the next few years, the FHD market should receive significant competition from both the solid-state technologies, such as CCD and magnetic bubbles, and the moving-head disk technology, which has been constantly improving relative to the fixed-head disk in terms of cost and performance. Because of the high cost per bit stored of FHDs, the major advances in disk technology will be in the direction of improved moving-head disk performance and the combination of solid-state and moving-head disk technology. Although the technology will move away from the FHD, it will probably still show at least an order of magnitude improvement in some areas of performance by 1985.

High-output thin-film alloys in the storage surfaces and thin-film read/write heads will be used to achieve higher areal densities in disk technology. The improved read/write heads will be designed to minimize the transverse sensitivity to avoid adjacent track pickup, which is a primary source of error, causing timing shifts in the desired data. However, because reduced signal-to-noise (S/N) ratio increases sensitivity to errors, the head electronics will be designed to maintain an adequate S/N ratio as the track width reduces and the bandwidth increases. Another source of error is head misregistration. These errors will be overcome with greater accuracy in head positioning and improved error correction codes.

In summary, it is the opinion of numerous experts in the field today that no fundamental limiting factors have yet emerged to limit future storage density advances and that the key areas for technology advances in FHD as well as MHD are minimization of track/head misregistration, maintenance of adequate S/N ratio at high densities without increasing sensitivity to errors, and reduction in head-disk spacing. Figure 7.2.4.1.3-1 presents the projected characteristic of fixed-head disks in the 1980-1985 timeframe.

TABLE 7.2.4.1.3-1. PROJECTED TECHNOLOGY FOR FIXED-HEAD DISKS, 1980-1985

SYSTEM CHARACTERISTICS	FIXED-HEAD DISKS
Track Density	1,000 tpi
Linear Density	10K bpi
Areal Density	Up to 10^7 bits/in^2
Drive Capacity	Up to 500 Mbytes
Maximum System Capacity	Up to 2.0 Gbytes
Rotation Time	4 to 6 msec
Average Access Time	2.5 to 12 msec
Transfer Rate	18.75 Mbytes/sec
Drive Cost	0.015 ¢/bit
System Cost	0.04 ¢/bit
MTBF	15,000 hr

7.2.4.2 Moving-Head Disks -

7.2.4.2.1 State of the Art in Moving-Head Disks - Over the past 12 to 15 years, disks have provided the bulk of on-line mass storage. Moving-head disk technology is the dominant disk storage technology today. A disk storage system is configured from disk drives and controllers. A moving-head disk drive consists of a stack of rigid magnetic disks addressed by positionable read-write heads that are mounted on arms attached to a voice coil actuator. The actuator moves the heads from track to track so that each head covers a large number of concentric tracks, thereby reducing the cost of access to large amounts of data since the read-write heads contribute a major portion of the overall cost. Until recently, the majority of the drives incorporated removable disk packs. However, non-removable packs are appearing in many of the newest drives. Table 7.2.4.2.1-1 summarizes the characteristics of state-of-the-art moving-head disk systems. The largest system capacities are associated with the largest spindle capacities, which in turn are associated with the largest track and linear densities. The drives with the largest spindle capacities generally allow the highest transfer rates. Because the larger spindle capacities are achieved by packing bits more closely together, the physical dimensions of systems are not proportional to their capacities. Some of the largest capacity systems occupy the least amount of space and have the smallest weight.

As a magnetic recording type of storage, moving-head disks have the advantages of nonvolatility; low cost; allocation flexibility; a simple, reliable recording process; and update in place. Their major disadvantage is that they involve mechanical motion and thus cannot achieve the speeds that all-electronic storage can achieve. In several of the newest disk systems, fixed-head storage is being combined with moving-head storage on the same spindle to reduce the average access time, albeit at an increased cost. This technique is particularly useful when multiple seeks are required such as for indexing on the disk.

TABLE 7.2.4.2.1-1. STATE OF THE ART IN MOVING-HEAD DISKS

SYSTEM CHARACTERISTICS	REMOVABLE PACK	NONREMOVABLE PACK
Track Density	192 to 402 tracks/in.	402 to 480 tracks/in.
Linear Density	Up to 6,425 bpi	Up to 6,425 bpi
Areal Density	Up to 3 Mbits/in.^{2}	Up to 3 Mbits/in.^{2}
Spindle Capacity	69.8 to 317.5 Mbytes	300 to 400 Mbytes
Maximum System Capacity	800 to 6,400 Mbytes	5,000 to 12,800 Mbytes
Typical Physical Characteristics of Minimum-Capacity System:		
Width by Depth by Height	19 by 34 by 38 to 44 by 32 by 47 in.	20 by 34 by 40 to 50 by 42 by 33 in.
Weight	480 to 1,100 lb	500 to 1,100 lb
Temperature Range	60 to 90°F	60 to 90°F
Humidity Tolerance	20 to 80%	20 to 80%
Head Movement:		
Minimum	7 or 10 msec	7 or 10 msec
Average	25 to 30 msec	25 or 27 msec
Maximum	50 or 55 msec	50 msec
Rotation Time	16.7 msec	16.7 msec
Average Access Time	34.3 to 38.3 msec	33.4 to 35.3 msec
Transfer Rate	806,000 or 885,000 or 1,198,000 bytes/sec	806,000 or 1,198,000 bytes/sec
Cost	0.001 ¢/bit	0.0007 ¢/bit
MTBF	NA	NA

Moving-head disk storage is used primarily as on-line bulk memory, providing large quantities of reliable, low-cost storage. The cost per bit of this storage is heavily dependent on the overall system capacity. The cost per bit figures given in Table 7.2.4.2.1-1 are for large-capacity systems, those with at least 100 Mbytes of storage. Cost per bit in small-capacity systems can be 0.04¢ and higher.

The achievement of higher storage densities has required improvement in magnetic recording resolution, which has been achieved through improvements in head gap, head-disk spacing, and medium thickness. Improved track densities have followed from the advent of the voice coil actuator and closed-loop track-following systems. Present track-following systems contain a servo head in a multiple-head array to supply feedback information on head positioning. Since all heads move together, if the servo head is maintained on its track, then all other heads will be on their tracks.

High recording densities and medium flaws have made the use of error-correcting codes a standard procedure in state-of-the-art disk systems for improved system reliability. The most common codes in use are cyclic codes that have burst error detection and correction capabilities. Relatively short codes in conjunction with logical means for skipping defective surface areas can provide the desired error-rate levels. Increased reliability has also been gained through the use of the Winchester design, in which the recording disks, moving arm, and read-write heads are contained in environmentally sealed data modules. The reliability advantages of this technique include reading of data by the same head that writes the data, low contamination by airborne particles, and light head loading.

7.2.4.2.2 Trends in Moving-Head Disks - Moving-head disk storage is still an evolving and viable technology. There is widespread belief that during the next decade the use of disk storage will expand and technical advances will continue at the same rate as in the past, providing further major improvements in most performance categories, with as much as an order of magnitude advance in many categories. Moving-head disk storage will continue to be the dominant technology for on-line bulk storage, at least into the 1980's.

Most of the improvements in the next decade will be in the form of increased areal density resulting from more tracks per inch laterally and more bits per inch longitudinally, as illustrated in Figures 7.2.4.2.2-1 through 7.2.4.2.2-3. Increased densities will be achieved through improvements in recording resolution, as a result of decreased medium thickness, head-gap and head-disk spacing, and through improved head performance. An areal density improvement factor of at least 40 appears theoretically possible. Although advances in track density will be less dramatic, further gains can be anticipated from improvements in head-positioning accuracy resulting from the incorporation of servo information within the data along each track, to be sensed by the same head that performs the reading and writing, to reduce all possible mechanical tolerances. In addition, disk systems will continue to be designed for ever larger capacity per arm, thereby reducing the access rate per unit of capacity. Increasing arm capacity has been the basis for the trend toward reduced cost per bit. The 20 to 25% annual reduction in cost per bit will continue through the next decade, as illustrated in Figure 7.2.4.2.2-4.

Removable disk technology became attractive as a way of keeping drive costs down while increasing system storage capacity. With today's large-capacity drives, the need to remove packs during daily operation has been minimized. Fixed-disk packs are becoming desirable once again to achieve the tighter mechanical tolerances needed to achieve ever higher densities at the least cost.

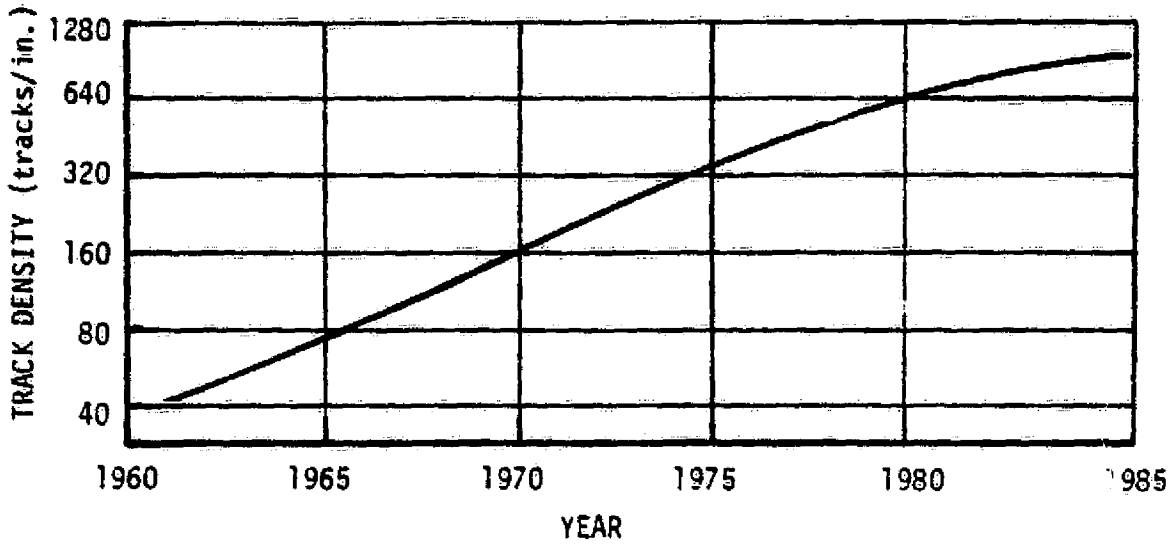


FIGURE 7.2.4.2.2-1 MOVING-HEAD DISK TRACK DENSITY TRENDS

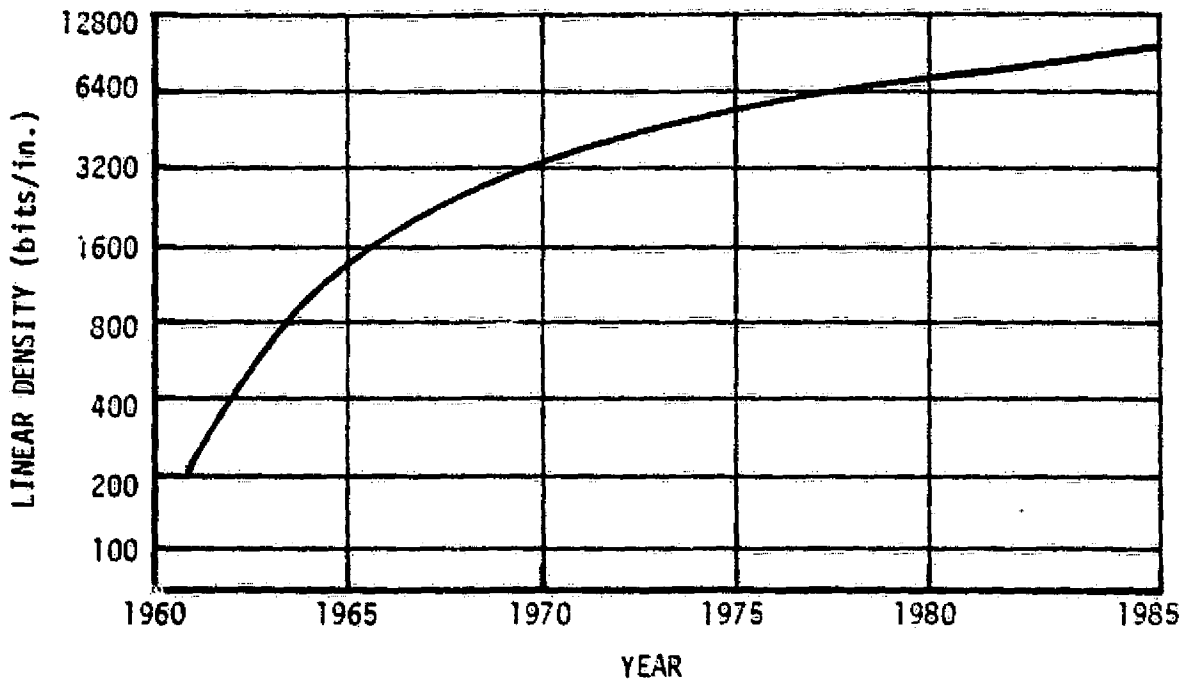


FIGURE 7.2.4.2.2-2. MOVING-HEAD DISK LINEAR DENSITY TRENDS

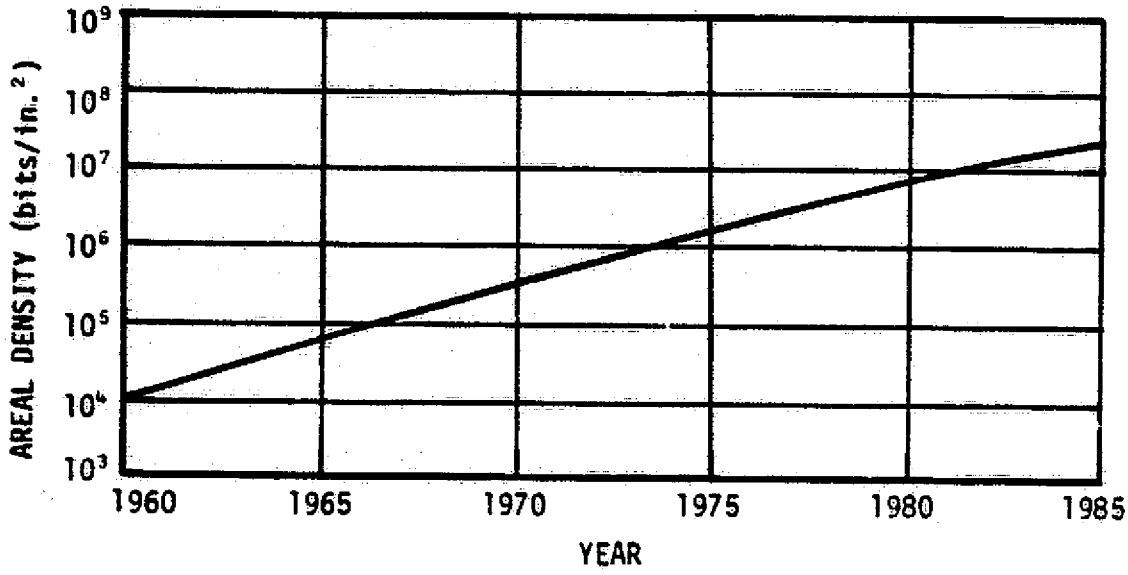


FIGURE 7.2.4.2.2-3. MOVING HEAD DISK AREAL DENSITY TRENDS

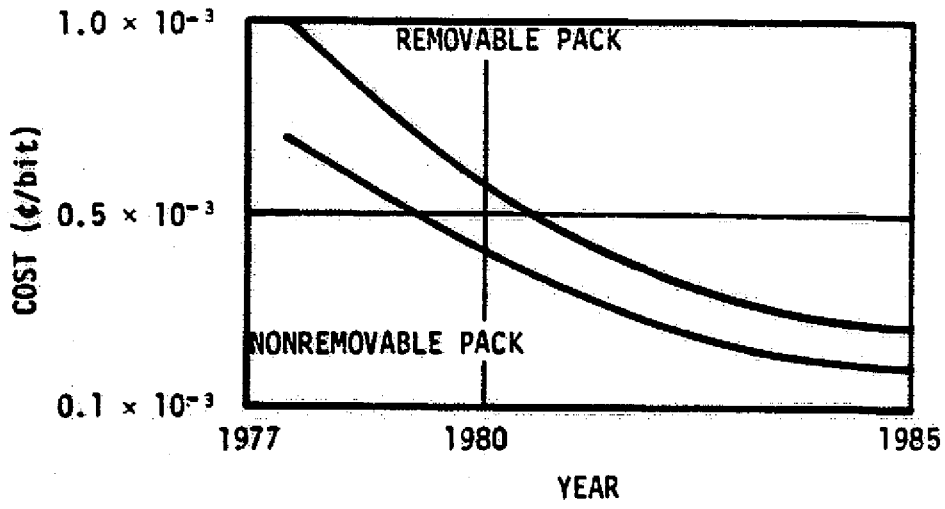


FIGURE 7.2.4.2.2-4. MOVING HEAD DISK COST TRENDS

7.2.4.2.3 Projected Developments in Moving-Head Disks - The trends in moving-head disk technology over the past 15 years and the assessments of several authorities in the field indicate that moving-head disk systems will be commercially available in 1980 and 1985 with the characteristics given in Tables 7.2.4.2.3-1 and 7.2.4.2.3-2, respectively. As discussed under trends (Subsection 7.2.4.2.2), the progress in storage densities will be achieved in the next decade through improved recording resolution. The advances will require a scaling down of all recording dimensions. Medium-base coatings will be thinner, with better dispersal and orientation of the particles. Read-write heads will be spaced more closely and will have improved performance, with minimized adjacent track pickup, minimized head misregistration, and increased longitudinal head resolution. Utilization of these storage density advances to increase the storage capacity per drive will cause a decrease in cost per bit and average access time without changing the physical characteristics of the system. The cost per bit of moving-head disk storage will continue to depend heavily on the overall system capacity. The cost per bit figures given in Tables 7.2.4.2.3-1 and 7.2.4.2.3-2 are for large-capacity systems.

Increasing positioner motor force constants by the use of rare-earth magnet motors will aid the development of faster access times during the next decade. In addition, effective positioner speed will be improved through the use of multitrack heads in some systems. They will be used in fixed-head files, as fixed-head additions to moving-head files, and as moving multitrack heads in a hybrid design. An added benefit of this approach is that it increases the size of a cylinder, thereby improving performance in the case of sequential operations.

Improvements in system reliability will be aided by several factors during the next decade. The use of error-correcting codes will expand since, as a rule of thumb, data integrity is improved by three orders of magnitude by the use of these codes. As electronic costs decline, error correcting codes will be built into the drives, making them transparent to the channel. Defect skipping will continue to be used, allowing the drive to locate magnetic surface flaws, to catalog them, and then, by the use of buffers, to skip over them during write and subsequent read operations. Redundancy

TABLE 7.2.4.2.3-1. PROJECTED TECHNOLOGY FOR MOVING-HEAD DISKS, 1980

SYSTEM CHARACTERISTICS	
Track Density	800 tracks/in
Linear Density	7,000 bits/in
Areal Density	9.4 Mbits/in^2
Spindle Capacity	600 Mbytes
Maximum System Capacity	19.2 Gbytes
Typical Physical Characteristics of Minimum-Capacity System:	
Width by Depth by Height	NA
Weight	NA
Temperature Range	60 to 90°F
Humidity Tolerance	20 to 80%
Head Movement:	
Minimum	6 to 10 msec
Average	23 msec
Maximum	45 msec
Rotation Time	16.7 msec
Average Access Time	30 msec
Transfer Rate	1.5 Mbytes/sec
Cost:	
Removable Pack	0.0005 ¢/bit
Nonremovable Pack	0.00036 ¢/bit
MTBF	NA

TABLE 7.2.4.2.3-2. PROJECTED TECHNOLOGY FOR MOVING-HEAD DISKS, 1985

SYSTEM CHARACTERISTICS	
Track Density	1,000 tracks/in
Linear Density	10,000 bits/in
Areal Density	25.4 Mbits/in^2
Spindle Capacity	1,000 Mbytes
Maximum System Capacity	32 Gbytes
Typical Physical Characteristics of Mimimum Capacity System:	
Width by Depth by Height	NA
Weight	NA
Temperature Range	60 to 90°F
Humidity Tolerance	20 to 80%
Head Movement:	
Minimum	5 to 8 msec
Average	20 msec
Maximum	40 msec
Rotation Time	16.7 msec
Average Access Time	27 msec
Transfer Rate	4.5 Mbytes/sec
Cost:	
Removable Pack	0.0002 ¢/bit
Nonremovable Pack	0.0001 ¢/bit
MTBF	NA

of electronics and tracks will be included in more systems. The use of fixed packs will expand, allowing greater control over possible damage caused by airborne particle contamination.

Moving-head disk technology is relatively mature. It should continue to be a reliable, cost-effective mass storage device during the 1980-1985 timeframe. Although other emerging technologies such as CCD and bubble memories will challenge some applications of these disks, their biggest threat during the next decade seems to be to the smaller disk systems that have a higher cost per bit.

7.2.4.3 Floppy Disks -

7.2.4.3.1 State of the Art in Floppy Disks - Floppy disks are the least expensive form of disk storage and offer a wide variety of design features. The floppy or flexible disk is a disk file storage medium that is removable and is enclosed in its own protective container (jacket). The floppy disk drive was developed because many applications could not tolerate the long access time of serial storage devices like cassette or cartridge tape drives or the high cost of large random-access disk drives.

The current floppy, or diskette, drive was introduced in 1973 for low-cost key entry applications using contract recording technology at surface speeds similar to half-inch tape drives. The newest generation of floppy disk drives have quadrupled storage capacity over earlier floppy drives by recording data on both sides of the diskette and recording at twice the standard density. The mini-floppy was developed to reduce the entry cost of a standard floppy disk system but retain high equipment reliability. This is an example of where available technology was applied to a developing need to create an entirely new end product. Another innovation is the dual-disk floppies that have two drives packaged in a single cabinet with a common controller and computer interface. Table 7.2.4.3.1-1 presents the state-of-the-art characteristics of full-sized and mini-floppy disk systems.

The floppy diskette consists of a 100-μin-thick layer of non-oriented magnetic oxide on a 0.003-in.-thick biaxially oriented polyester substrate (mylar), all enclosed in a protective jacket. This jacket has holes in it to provide head and index hole access and drive spindle mounting. The newer floppy disk drives use a head of the flying type that is separated from the media by about 20 μin. The older style in-contact heads were composed of a stainless steel body with a laminated permalloy read/write core. The newest heads use a glass-bonded ceramic design that has a ferrite magnetic core and a titanite face to improve

TABLE 7.2.4.3.1-1. STATE OF THE ART IN FLOPPY DISKS

SYSTEM CHARACTERISTICS	FULL-SIZED FLOPPY*	MINI-FLOPPY*
Track Density	48 tpi	48 tpi
Linear Density**	3,200 to 6,816 bpi	2,580/5,160 bpi
Unformatted Capacity	400K to 1,600K bytes	110K/220K bytes
Number of Tracks	77	35
Transfer Rate	250K/500K bps	125K/250K bps
Relative Velocity of Medium Over Head	120 ips	80 ips
Spindle Speed	360 rpm	300 rpm
Average Access Time	91 msec	566 msec
Drive Size	4.6 by 8.5 by 14.2, in.	4 by 6 by 8 in.
Weight	10 to 20 lb	3 to 15 lb
Power Consumption	72W	15W
Drive Price Less Controller	$600	$390
Media Size	8 by 8 in.	5.25 by 5.25 in.
Media Price†	$6.50/diskette	$4.50/diskette
Operating Temperature	10 to 45°C	10 to 45°C
Humidity Tolerance	20 to 80%	20 to 80%
Soft Error Rate	10^{-9}	10^{-8} to 10^{-9}
Hard Error Rate	10^{-12}	10^{-12}
Cost: Drive	5 m¢/bit	22 m¢/bit
Cost: System	7 to 90 m¢/bit	33 to 100 m¢/bit
MTBF	8,500 hr	8,500 hr

*Anywhere a / is used, it represents single-density floppy/double-density floppy.

**Inner-track density.

†Currently there are three types of noninterchangeable diskette media: IBM soft-sectored, hard-sectored, and Memorex. These differ mainly in the way tracks are sectored for storage and in the hole arrangements in the jacket.

head and media life. These newer head designs also use a tunnel erase structure borrowed from disk-head technology to clean up the track edges when overwriting old data.

In the last years the floppy disk has been successfully competing for the market and application areas of tape cartridge and cassette drives. The key implication of these developments is that magnetic disks continue to be an evolving and viable low-cost technology.

7.2.4.3.2 Trends in Floppy Disk - During the next decade the use of low-cost disk storage will expand and technical advances will closely follow those of the high-performance, large-capacity disk drives, providing further improvements in the following performance areas:

- Larger storage capacity
- Increased areal density
- Improved disk media
- Use of soft sectoring instead of hard sectoring
- Faster access
- Higher transfer rates
- Higher MTBF and lower MTTR.

As the floppy disk drive's performance increases, so will its variety of applications. By 1980, the floppy will be considered a low-cost, direct-access storage device in addition to its role for data entry and its load/dump function, while the mini-floppy will be used in desk-top computer applications. Even though the cost per bit of storage is higher for mini-floppies than low-cost tape drives, their random access, small size, and reliability will make them competitive in small-scale systems applications. Improved access time will result from the use of a new band-type head-positioning mechanism instead of using the conventional load screw drive mechanism. Figures 7.2.4.3.2-1 and 7.2.4.3.2-2 present cost and capacity trends for floppy disk drives. Figure 7.2.4.3.2-1 compares the cost trend of bubble memories and floppy disks through 1985. By 1985, the bubble memory, as on-line storage, can be expected to inhibit floppy disk growth, especially in the mini/micro market.

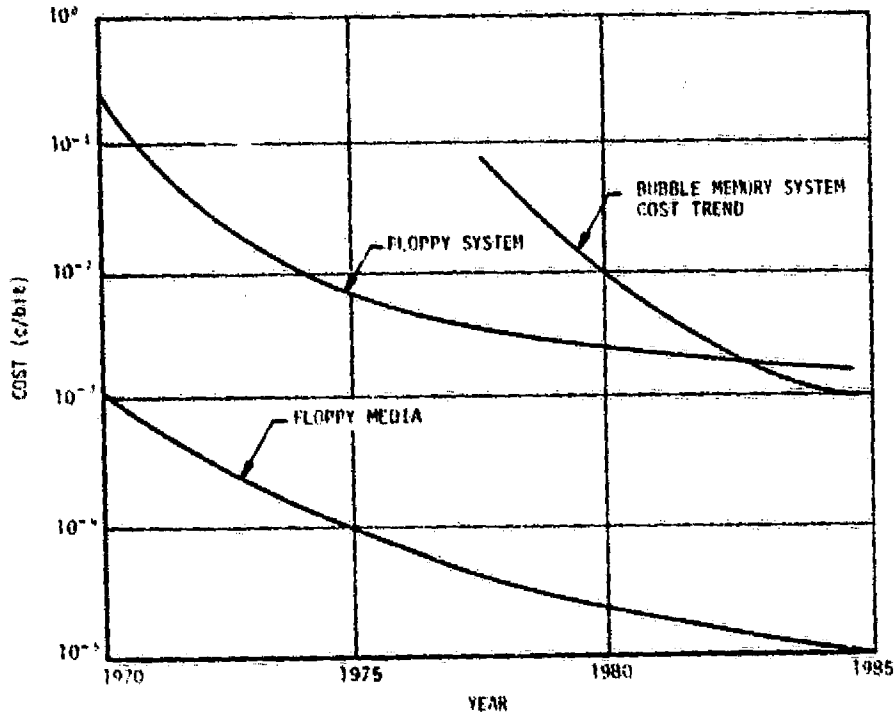


FIGURE 7.2.4.3.2-1. TRENDS IN FLOPPY DISK COST

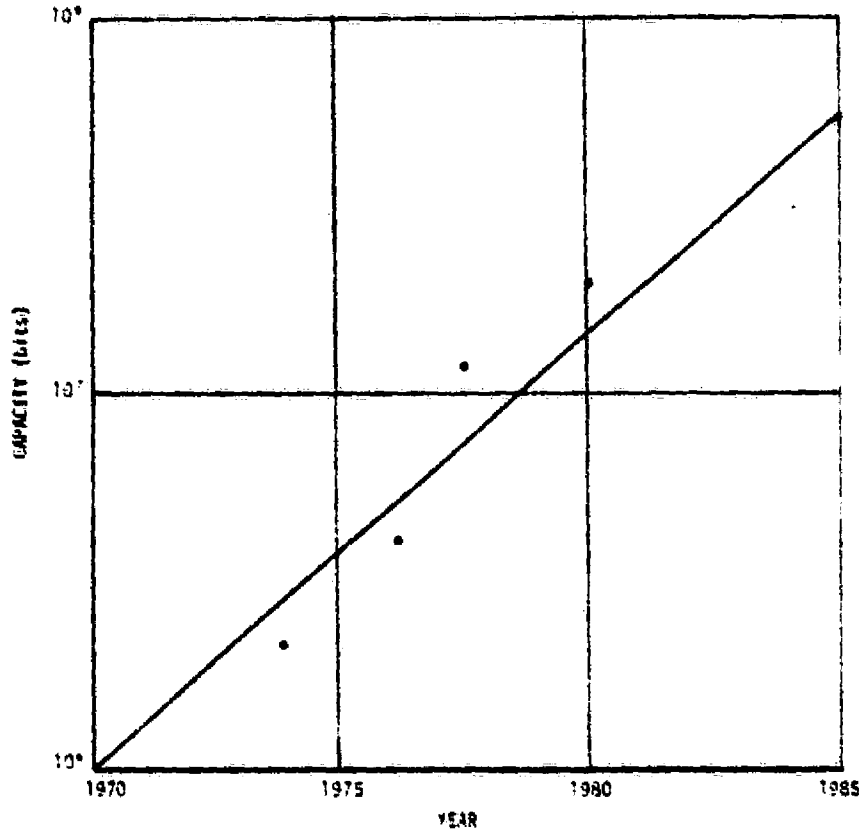


FIGURE 7.2.4.3.2-2. TRENDS IN FLOPPY DISK STORAGE CAPACITY

7.2.4.3.3 Projected Developments in Floppy Disks - The trends in floppy disk technology in the last 5 years and the assessment of several authorities in the field indicate that floppy disk drives will be a viable product through the early 1980's, safe from the advancing technology of solid-states storage devices such as bubble memories. The projected characteristics of full-sized and mini-floppy disk systems for 1985 are presented in Table 7.2.4.3.3-1.

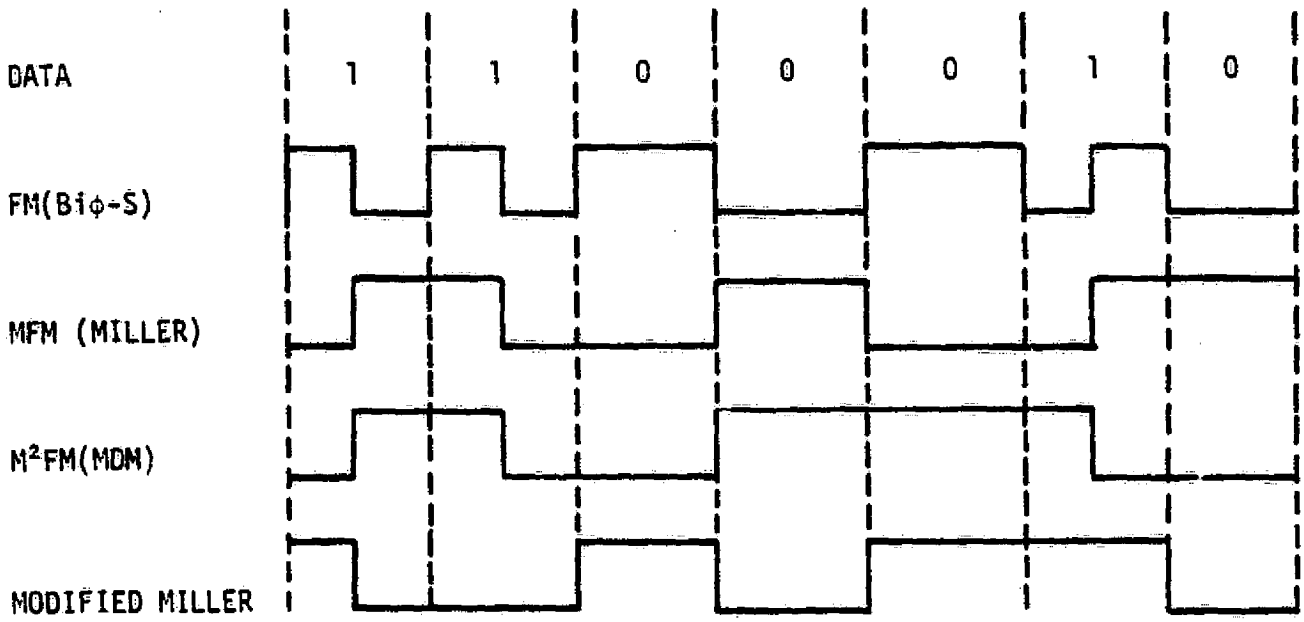
By 1980, floppy storage capacity will increase to 2.4 Mbytes by doubling the current 77 recording tracks on a diskette surface to 154 tracks. As much as an order of magnitude reduction in the price of floppy disk systems seems feasible if high-volume usage continues to develop. The main problem with future floppies and other low-cost electromechanical systems will be reliability. If maintenance costs for floppy disk systems are high, industry will probably turn to all-electronic storage devices such as bubble memories, even if acquisition costs appear to be somewhat higher. Closed-loop actuation with track-following servo control promises even further enhancements in storage capacity without significantly altering the media coating or configuration, thus allowing diskette interchangeability. The stepping motors currently used for head positioning will be replaced by more reliable linear motors similar to the ones currently used in moving-head disk drives. Improved throughput and faster access rates will come about with the use of metal foil substrates instead of mylar, flat read/write heads, and increased rotation speeds. Increased storage capacity will be attained with the use of multi-disk recording with two or more diskettes on the same spindle.

Another area in which improvement is expected is in the use of more efficient data encoding techniques. Most floppies currently use frequency modulation (FM) encoding. Three double-density encoding methods are: modified frequency modulation (MFM), also called Miller code; modified-modified frequency modulation (M^2FM) codes, also called modified-delay-modulation code; and modified Miller code. These codes are illustrated and explained in Figure 7.2.4.3.3-1.

TABLE 7.2.4.3.3-1. PROJECTED DEVELOPMENTS IN FLOPPY DISKS, 1985

SYSTEM CHARACTERISTICS	FULL-SIZED FLOPPY	MINI-FLOPPY
Track Density	144 tpi	96 tpi
Linear Density*	9,600 bpi	8,000 bpi
Unformatted Capacity	6,000K bytes	800K bytes
Number of Tracks	231	77
Transfer Rate	880K bps	440K bps
Relative Velocity of Medium Over Head	120 ips	80 ips
Spindle Speed	360 rpm	300 rpm
Average Access Time	80 msec	300 msec
Drive Size	NA	NA
Weight	10 to 20 lb	3 to 5 lb
Power Consumption	60W	15W
Drive Price Less Controller	400	195
Media Size	8 by 8 in.	5.25 by 5.25 in.
Media Price	$3.20/diskette	$2.00/diskette
Operating Temperature	NA	NA
Humidity Tolerance	NA	NA
Soft Error Rate	10^{-10}	10^{-10}
Hard Error Rate	10^{-12}	10^{-12}
Cost: Drive	1 m¢/bit	3 m¢/bit
Cost: System	2 m¢/bit	5 m¢/bit
MTBF	10,000 hr	10,000 hr

*Inner-track density.



ENCODING RULES

FM - 1) WRITE "1" BIT AT CENTER OF THE BIT CELL.
2) WRITE CLOCK BIT AT THE LEADING EDGE OF THE BIT CELL.

MFM - 1) WRITE "1" BIT AT CENTER OF THE BIT CELL.
2) WRITE CLOCK BIT AT THE LEADING EDGE OF BIT CELL IF:
a) NO "1" HAS BEEN WRITTEN IN THE PREVIOUS BIT CELL
b) NO "1" WILL BE WRITTEN IN THE PRESENT BIT CELL.

M^2FM - 1) WRITE "1" BIT AT THE CENTER OF THE BIT CELL.
2) WRITE CLOCK BIT AT THE LEADING EDGE OF THE BIT CELL IF:
a) NO "1" OR CLOCK BIT HAS BEEN WRITTEN IN THE PREVIOUS BIT CELL
b) NO "1" BIT WILL BE WRITTEN IN THE PRESENT BIT CELL.

MODIFIED MILLER - 1) WRITE "0" AS A CLOCK AT LEADING EDGE OF BIT CELL.
2) WRITE "1" AS EITHER BIT AT CENTER OF BIT CELL OR NO TRANSITION DEPENDING ON DC COMPONENT OF DATA.

FIGURE 7.2.4.3.3-1. DESCRIPTION OF FM, MFM, M^2FM, AND MODIFIED MILLER ENCODING TECHNIQUES

7.2.4.4 Cartridge Disks -

7.2.4.4.1 State of the Art in Cartridge Disks - Disk cartridge drives are used in computer systems today to fill the cost/performance gap between floppies and the larger removable disk packs. Cartridge disk drives cost about three times more than floppy drives but provide an order of magnitude better performance. The typical cartridge drive has an average access time of 55 msec, a storage capacity up to 13 Mbytes, and a transfer rate of 187 Kbytes/sec. Disk cartridges come in two versions: the top-loading IBM-standard 5440 cartridge type and the front-loading IBM-standard 2315 cartridge type. Disk cartridge drives are available in several configurations, including: one fixed, one removable, or one removable combined with one to three fixed disks. The concept of combined fixed and removable disk cartridges was developed to fill the need for more storage capacity, file backup, and copying without constantly changing drive mechanism and recording medium. Table 7.2.4.4.1-1 summarizes the characteristics of the single removable disk and the combination of removable disk with up to three fixed-disk platters.

TABLE 7.2.4.4.1-1. STATE OF THE ART IN DISK CARTRIDGE SYSTEMS

SYSTEM CHARACTERISTICS	REMOVABLE DISKS	COMBINATION OF FIXED AND REMOVABLE DISKS
Track Density	100 to 200 tpi	100 to 500 tpi
Linear Density	1,100 to 2,200 bpi	2,200 to 4,750 bpi
Drive Capacity	2.5 to 13.3 Mbytes	5 to 53.3 Mbytes
Transfer Rate	187K to 312K bps	Up to 937.5K bps
Rotational Speed	1,500 to 2,400 rpm	1,500 to 3,600 rpm
Track-to-Track Access Time	6 to 15 msec	6 to 15 msec
Average Access Time	35 to 70 msec	8.3 to 45 msec
Maximum Access Time	60 to 80 msec	55 msec
Soft Error Rate	10^{-10}	10^{-10}
Hard Error Rate	10^{-12}	10^{-12}
Single Drive Price	$2,600 to $5,000	$3,500 to $6,500
Controller Price	$3,000 to $5,000	$3,000 to $6,000
Disk Cartridge Price	$68 to $180	$68 to $180
Temperature Range	55 to 90°F	55 to 90°F
Humidity Tolerance	20 to 80%	20 to 80%
System Cost	5 to 10 m¢/bit	2 to 4 m¢/bit
Media Cost	0.1 m¢/bit	0.1 m¢/bit
MTBF	NA	NA

7.2.4.4.2 Trends in Cartridge Disks - Current trends in cartridge disk drives include: more cartridges per drive, more fixed disks on the same spindle, and increased capacity per cartridge. During the next decade, the increased need for more operating and backup storage will result in cartridge disk drives with between one and three removable cartridges and five to six fixed disks on the same spindle. Areal density improvements will provide storage capacities of up to 100 Mbytes by 1985.

Another trend is the use of higher rotation speeds to decrease access time and increase data transfer rates. These trends and improvements will place disk cartridge drives in performance competition with the standard disk pack systems. To produce competitive cost/performance ratios and maintain reliability, the new improvements will require higher packing densities at low cost, improved head design, and more precise head positioning. One of the newest cartridge drives, the Diablo series 400, is moving away from previous cartridge drive technology. This drive uses a flywheel principle of operation for two actuators (one for the removable disk and the other for the fixed disks), an inertial power supply, direct track positioning, and a dual-microprocessor-based controller. The inertial actuator eliminates the usual large magnetic fields and power requirements normally associated with traditional voice coil positioners. Other trends include the use of microprocessors for error detection and correction, rotational position sensing, alternate track switching, and track servo technology.

7.2.4.4.3 Projected Developments in Cartridge Disks - The trends in disk cartridge technology, in the opinion of several experts in the field, are toward substantial progress in disk cartridge drive performance by 1985 because of the strong commercial demand for inexpensive mass storage devices to serve low-cost processing units made possible by LSI technology. Table 7.2.4.4.3-1 contains characteristics of commercially available disk cartridge drives between 1980 and 1985. Because technology for these increased performance characteristics already exists or is very near to development, no major breakthrough is needed. A previously discussed projected development is the use of multiple fixed disks on the same spindle with the removable disk to provide users with cheaper storage, more capacity, and greater operational convenience. Each additional fixed disk will allow increases in storage capacity from 5 to 20 Mbytes. This feature will also allow electronic track switching up and down the cylinder, thus sharply reducing time delays due to moving head seeks. Another important projected development is the use of multiple fixed heads on each of the added fixed disks to enhance programs and file swapping performance.

TABLE 7.2.4.4.3-1. PROJECTED DEVELOPMENTS IN DISK CARTRIDGE SYSTEMS, 1980 - 1985

SYSTEM CHARACTERISTICS	1985
Track Density	800 tpi
Linear Density	6,600 bpi
Drive Capacity	120 Mbytes
Transfer Rate	Up to 1,200 Kbytes/sec
Rotational Speed	Up to 3,600 rpm
Track-To-Track Access Time	3 msec
Average Access Time	25 msec
Maximum Access Time	50 msec
Soft Error Rate	NA
Hard Error Rate	NA
Single Drive Price	$2,000 to $6,000
Controller Price	$1,500 to $4,000
Disk Cartridge Price	$80
Temperature Range	NA
Humidity Tolerance	NA
System Cost	1 to 6 m¢/bit
Media Cost	0.09 m¢/bit
MTBF	NA

7.2.5 Magnetic Tape Systems

7.2.5.1 Half-Inch Reel-To-Reel Tape

7.2.5.1.1 State of the Art in Half-Inch Reel-to-Reel Tape - Half-inch reel-to-reel magnetic tape is the major medium used for storage and interchange of formatted digital data throughout the world. Since the first commercial reel-to-reel tape drive in 1953, the recording density has increased from 100 to the present 6,250 bpi, and throughput rates have increased from 15,000 to 1,250,000 bytes/sec. The standard configurations include both 7- and 9-track systems, with 9-tracks being predominant in newer systems. Reliable performance and low cost per bit of storage have made half-inch tape the most popular low-cost storage medium today, with more than 150,000 units currently in use.

Table 7.2.5.1.1-1 summarizes the state-of-the-art characteristics for low-cost and high-performance reel-to-reel tape drives. The high-performance drives are generally characterized by much faster start/stop times, lower error rates, higher density recording, and mechanical features that facilitate tape/mounting and dismounting.

A major advantage of half-inch tape drives is the upward/downward compatibility of tapes. Tapes written 20 years ago can still be read by today's tape drives with little if any modifications. Other advantages for on-line secondary storage are: high capacity, low cost per bit stored, high transfer rates, archival stability, relatively simple and reliable read/write electronic techniques, and the ability to selectively change the data and thereby reuse the recording surface. The disadvantages of tape drives in general are: slow access times, high cost for labor in terms of tape mounting and library maintenance, and sequential access.

Today's reel-to-reel tape systems are characterized by the use of microprocessors and microprogrammed circuitry within drives, controllers, and interfaces that permit significant flexibility in throughput, error detection and correction, and interfacing. Features available in current drives include microprogrammed diagnostics, dual-density recording, automatic threading, dynamic amplitude and braking control, velocity feedback

TABLE 7.2.5.1.1-1. STATE OF THE ART IN HALF-INCH REEL-TO-REEL TAPE DRIVES

SYSTEM CHARACTERISTICS	LOW-COST	HIGH-PERFORMANCE
Number of tracks	7 or 9	7 or 9
Linear Density	Up to 1,600 bpi	Up to 6,250** bpi
Operating Tape Speed	Up to 75 ips	Up to 250 ips
Maximum Transfer Rate	Up to 120 Kbytes/sec	Up to 1,250 Kbytes/sec
Tape Drives per Controller	Up to 4	Up to 16
Design Configuration	Serial	Radial
Simultaneous Read/Write	No	Yes
Rewind Time (for 2,400 ft)	60 to 240 sec	45 to 76 sec
Maximum Unformatted Capacity*	4.6×10^7 bytes/drive	1.8×10^8 bytes/drive
Typical Start/Stop Time (msec)	5 to 30	0.95 to 1.1
Recording Format	PE/NRZI	PE/GCR
Interrecord Gap Size (in.)	0.6 to 0.75	0.3
Media cost (2,400-ft reel)	0.03 m¢/bit ($8 to $12)	0.01 m¢/bit ($9 to $15)
System Cost	2 to 4 m¢/bit	2 to 4 m¢/bit
Drive Cost	$2.5K to $5.0K	Up to $38K
Controller Cost	$3.0K to $5.0K	Up to $50K
Drive Size (W by D by H)	19 by 24 by 15 in	N/A
Reel Size (Diameter)	7.0/8.5/10.5 in	7.0/8.5/10.5 in
Soft Error Rate	10^{-8}	10^{-9}

*No allowance for interrecord gap. For formatted capacity, refer to Table 7.2.5.1.1-2. Tape length was 2,400 ft.

**6,250 bpi is user density; tape density using GCR coding is 9,042 bpi.

reel control, automatic reel hub, analog capstan control, vacuum tape buffers, prealigned fixed-head assemblies that minimize head-to-tape alignment errors, linear sensors and reel servos, backward reading, and radial attachments to the controllers.

The use of advanced electronics and improved recording techniques has considerably increased the error detection and correction capabilities of today's tape drives. The 1,600/6,250-bit/in tape drives use a relatively new recording method called Group-Coded Recording (GCR), whereas the older drives used phase-encoding (PE) and nonreturn-to-zero inverted (NRZI) techniques. The GCR encoding method translates a 4-bit data subgroup into a 5-bit storage subgroup in such a manner that no more than two zeros follow each other for any combination of 5-bit storage groups. The 5-bit storage groups are then written on tape in the NRZ-M format.

Tape media have significantly improved since 1965. Improvements have taken place in the area of physical, chemical, and mechanical properties of the tape. Many of the newer tapes use iron-cobalt (Fe-Co) metallic particle coating to improve signal-to-noise (S/N) ratios and drop-in and drop-out rates. The terms drop-in and drop-out mean the number of bits gained or lost in a group of data bits and is typically 1 bit in 10^7 for drop-out rates. Even with improved tape capabilities, the high performance of 6,250-bits/in drives puts excessive demands on current tape quality and requires a large overhead in error correcting. At present, industry has found that the 3,200- to 4,000-bit/in range is the most cost-efficient operating density. This density range remains practical because it is consistent with current tape quality and requires only simple electronics for its recording and reproducing techniques.

The on-line system storage cost per bit for a 16-tape drive configuration with 6,250-bit/in capability is on the order of 2 to 4 mc/bit, assuming approximately 10^9-bit capacity per drive. System cost per bit also falls within the range of 2 to 4 mc for the low-cost systems. Of

course the cost per bit is much lower for the high-performance systems when considering the off-line storage capacity available. Table 7.2.5.1.1-2 presents the performance and storage factors for different block sizes and packing densities and tells how to use these factors to determine the effective transfer rate and storage capacity in bytes for a given system.

TABLE 7.2.5.1.1-2. TAPE DRIVE PERFORMANCE AND STORAGE FACTORS

RECORD BLOCK SIZE (bytes)	bits/in/INTERRECORD GAP		
	800/0.6	1,600/0.6	6,250/0.3
80	114.4	123.0	255.6
500	408.0	547.2	1,318.8
1,000	540.8	816.0	2,175.0
2,000	644.8	1,081.6	3,225.0
4,000	714.4	1,289.6	4,256.3
8,000	754.7	1,428.6	5,063.3

Effective transfer rate in bytes/sec = Appropriate factor × Tape Speed (ips)

Storage Capacity in bytes = Appropriate factor × Tape length (in.)

7.2.5.1.2 Trends In Half-Inch Reel-to-Reel Tapes - The trends in half-inch reel-to-reel drives are toward higher linear density, higher throughput, more efficient interfacing and utilization of the I/O channel, better quality tape, and automated tape operations. Figure 7.2.5.1.2-1 illustrates the trend in linear recording density. During the past decade, the linear recording density has almost doubled every 4 years. This trend will slow somewhat during the next decade since the increased costs of higher-density tapes drives are not offset completely by reduced media costs. However, evolutionary improvements in coding, media, and head designs should provide at least a two-fold additional improvement in reel-to-reel storage density by 1985. Even with a two-fold increase, the user areal storage density for a 9-track tape will be only 2×10^5 bits/in^2. Other types of digital tape recorders are currently under development that have user data storage densities in excess of 6×10^6 bits/in^2. It is therefore not unreasonable to assume that IBM, who traditionally establishes reel-to-reel tape standards, may introduce some type of high-performance (high track density) tape systems that offers at least an order of magnitude improvement in storage density and data transfer rate. Although the technology currently exists to develop such a system, other economic factors such as upward/downward tape compatibility with existing systems may preclude development.

Data transfer rate trends and system cost trends are shown in Figures 7.2.5.1.2-2 and 7.2.5.1.2-3. Data transfer rate and media costs are directly related to linear storage density. System cost reductions will result primarily from increased use of microminiature electronic circuitry. System cost decreases will be limited by application of more sophisticated coding techniques, improved head designs, etc. Further reduction in media costs (cost/bit) will result from increased storage density rather than lower cost tape manufacturing processes.

The current feeling of the experts in the industry is that half-inch magnetic tape technology will continue to dominate the off-line computer storage through 1985. Despite all the technological advances in magnetic tape recording, the major disadvantage, high labor cost, will continue through 1985 unless there is greater development in the use of automated file handling.

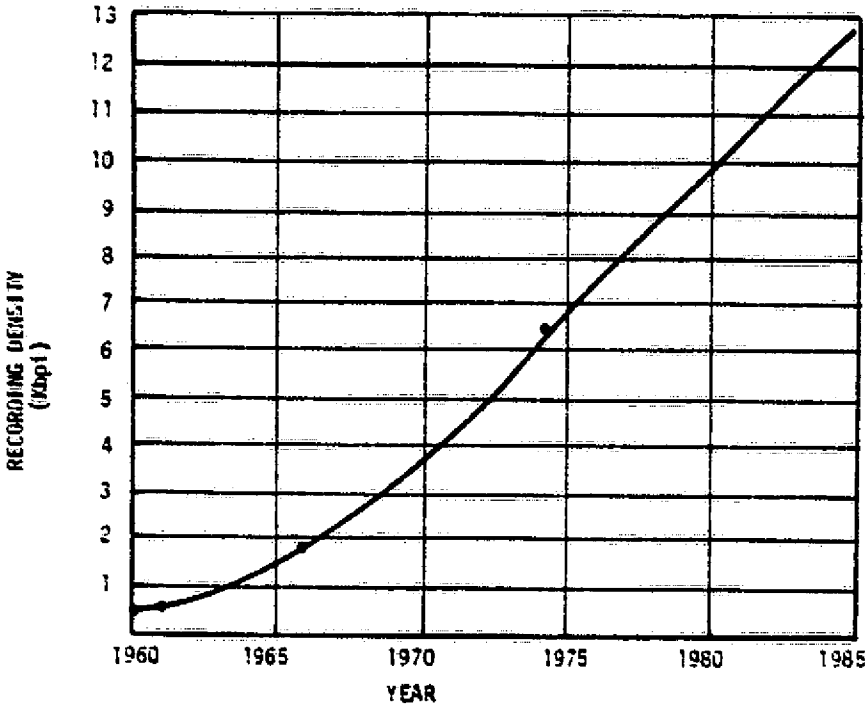


FIGURE 7.2.5.1.2-1. TRENDS IN REEL-TO-REEL TAPE LINEAR DENSITY

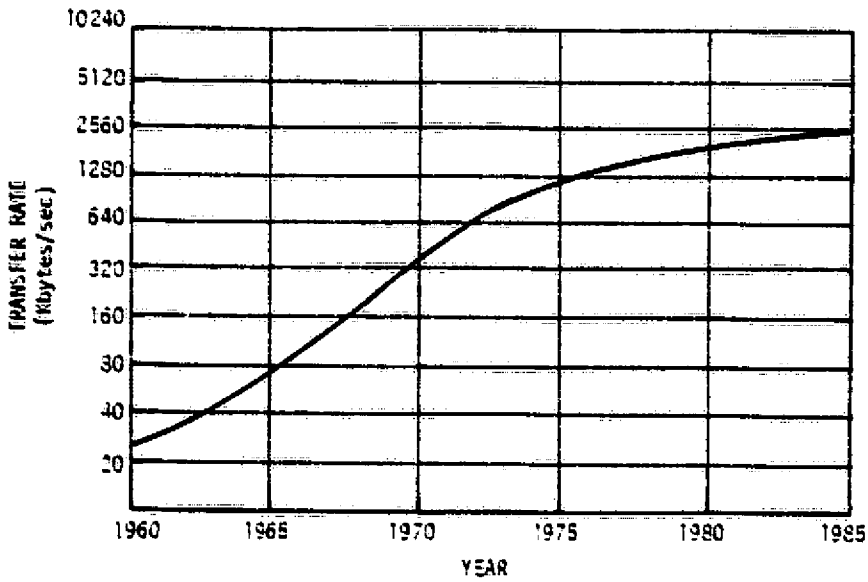


FIGURE 7.2.5.1.2-2. TRENDS IN REEL-TO-REEL TAPE TRANSFER RATE

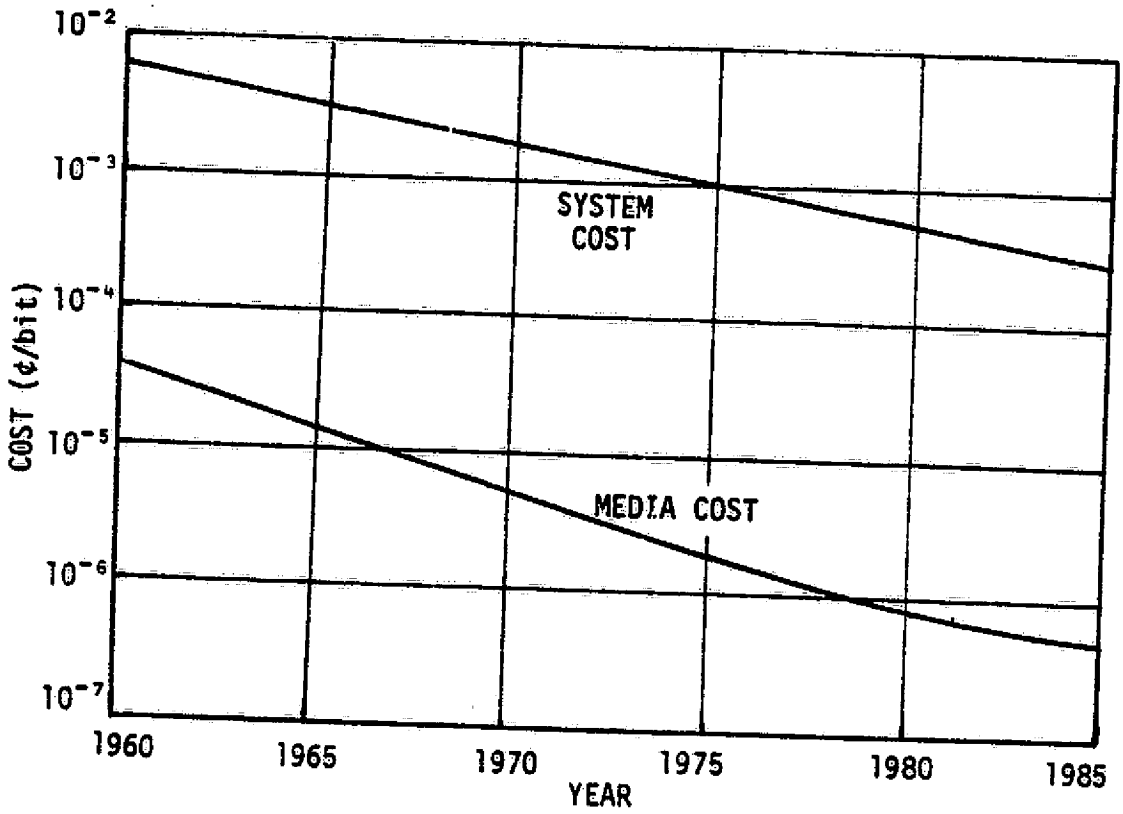


FIGURE 7.2.5.1.2-3. TRENDS IN REEL-TO-REEL TAPE COST

7.2.5.1.3 Projected Developments in Half-inch Reel-to-Reel Tapes - The trends in magnetic reel-to-reel tape technology over the last 25 years and the assessment of several authorities in the field indicate that half-inch tape drives will be a viable commercial product in 1980 and 1985 with the characteristics given in Tables 7.2.5.1.3-1 and 7.2.5.1.3-2, respectively. Trends in half-inch magnetic tape show that the future holds higher linear densities and transfer rates, lower cost, automated tape handling, improved reliability and accuracy, and modularity for faster on-line repair and maintenance.

Because the computer industry has standardized on 7- and 9-track half-inch reel-to-reel tape systems, improvements in recording density are limited to increasing the number of bits per inch of tape. Currently, much higher areal densities are being realized in digital instrumentation recorders (see Section 3.1) by utilizing a combination of higher track densities and higher bit densities. Although the technology currently exists to build much higher performance half-inch reel-to-reel tape systems, the existence of the 7- and 9-track standards and the already extremely low cost of the magnetic tape media will probably preclude the adoption of a new reel-to-reel standard during the next decade. One possible alternative to standard half-inch tape would be a small tape system based on the IBM 3850/CDC 38500 mass storage systems technology that includes automated file handling, high-density helical scan recording, or air-bearing heads. Another possible improvement is the use of air-bearing rollers, which allow the tapes to travel on a pillar of air such that tape wear is eliminated.

In addition to the system design, improvements will be made in tape media formulations to increase both the magnetic resolution and stability of the tape media. Computer tapes using finer metallic particles will probably be available by 1980. Also projected are developments to improve the tape viscoelastic material to reduce interlayer slippage of a tape when re-read after storage. Even though digital magnetic tape recording is a mature technology, the next decade will see this technology expanding with technical advances providing improved cost and performance.

TABLE 7.2.5.1.3-1. PROJECTED TECHNOLOGY FOR HALF-INCH REEL-TO-REEL TAPE DRIVES, 1980

SYSTEM CHARACTERISTICS	LOW-COST	HIGH-PERFORMANCE
Number of Tracks	7 or 9	9
Linear Density*	Up to 1,600 bpi	Up to 9,600
Operating Tape Speed	Up to 125 ips	Up to 250 ips
Maximum Transfer Rate	Up to 200K bytes/sec	Up to 1,900K bytes/sec
Tape Drives per Controller	Up to 8	NA
Maximum Unformatted Capacity**	4.6×10^7 bytes/drive	2.8×10^8 bytes/drive
Typical Start/Stop Time	4.4 to 20 msec	0.6 to 1.0 msec
Recording Format	NRZI/PE/GCR	GCR
Cost	3 mc/bit	0.95 mc/bit
Soft Error Rate	10^{-9}	10^{-11}

*User density

**No allowance for interrecord gap and 2,400 ft of tape.

TABLE 7.2.5.1.3-2. PROJECTED TECHNOLOGY FOR HALF-INCH REEL TAPE DRIVES, 1985

SYSTEM CHARACTERISTICS	LOW-COST	HIGH-PERFORMANCE
Number of Tracks	7 or 9	9
Linear Density*	Up to 3,200 bpi	Up to 12,800 bpi
Operating Tape Speed	Up to 200 ips	Up to 250 ips
Maximum Transfer Rate	Up to 640K bytes/sec	Up to 2,500K bytes/sec
Tape Drive per Controller	Up to 12	Up to 24
Maximum Unformatted Capacity**	9.22×10^7 bytes/drive	3.7×10^8 bytes/drive
Typical Start/Stop Time	2.0 to 20 msec	0.5 msec
Recording Format	NRZI/PE/GCR	NA
Cost	2 mc/bit	0.6 mc/bit
Soft Error Rate	10^{-9}	10^{-11}

*User density

**No allowance for interrecord gap and 2,400 ft of tape.

7.2.5.2 Cassettes and Cartridges - The cassette drives were originally developed from the low-cost Philips type cassette that is used for audio application, with modifications and improvements for digital recording. Today's cassettes are used where storage requirements and access time are relatively low and cost is a primary consideration. The cartridge drive technology is derived from a combination of reel-to-reel and cassette technology and provides medium storage capacity at a relatively low cost. In addition to standard cassette and cartridge drive configurations, many manufacturers are offering cassette/cartridge drives in miniature versions.

The cassette/cartridge drives are generally used in small terminal systems, data collection systems, data entry systems, and small business systems where low cost is an important factor. They are the most widely used medium since paper tape and punched cards, with over 600,000 operational units currently and a growth rate of over 150,000 additional units per year. The general difference between cassette and cartridge tape units has usually been that cartridge drives were larger in size, tape width, and tape length. Recently, the difference has changed with the innovation by 3M of smaller version cartridge drives.

7.2.5.2.1 State of the Art in Cassettes and Cartridges - Although the cassette drive has changed very little during the last 5 years, some improvements have been made on reliability, operating speed, packing density, and standardization of recording formats (ECMA or ANSI standards). The standard control functions of the cassette are: write, stop, reverse, read one block, read continuously, check read one block, erase, rewind to BOT, and rewind for cassette removal. Several data flow checking schemes currently on cassette drives are: read-after-write (RAW), drop-in and drop-out check, bit timing checks, cycle redundancy check (CRC), and data parity check. There are basically two types of digital cassette media: the Philips and the quarter-inch version. The quarter-inch cassette is not as popular as the Philips type because of its increased complexity and higher cost, although it has advantages over the Philips cassette in the areas of greater storage capacity and the ability to write four tracks in parallel. Cassettes used either the reel-to-reel or continuous-loop tape design. The continuous or endless loop design generally reduces tape wear and cost by eliminating the mechanical components required to rewind, engage, and drive reels, since the tape only moves forward for read/write passes. Most cassettes use a high-inertia capstan system that rotates all the time and therefore uses a "banging" pressure roller to move the tape. This "banging" is relatively hard on the tape and reduces the media reliability. Some of the major advantages of cassettes are that they are relatively insensitive to varying ambient conditions and are of simple and rugged construction, thus making them suited for a wide variety of applications.

The advantages of the cartridge tape systems are low cost, simple mechanics, compactness, low power consumption, high storage capacity, ruggedness, and long medium life. Cartridges generally use precision die-cast housing construction, ball bearings, and simple two-guide design to provide uniform tape transfer between reels. These design features reduce wear and flutter and improve data reliability. The simplicity of the design of tape cartridges eliminates some of the tape handling problems inherent in conventional tape transport designs and provides a totally

enclosed environment for rugged applications. The cartridge tape drive was developed because of the need for higher operating speeds and greater storage capacity than were available from cassette drives. Some approaches to increasing the data capability of a cartridge unit are the carousel system, which uses 16 quarter-inch tape cartridges in a removable pack to obtain 32 Mbytes of storage, or the use of double-density encoding techniques and phase-locked loop data separators to achieve 3,200 bits/in (double ANSI density).

One of the major disadvantages of cassette/cartridge systems is data reliability or bit error rate (BER). The errors that occur most often in cassette/cartridge tape systems are a result of excessive filtering and drive speed variations. These errors can be overcome by the use of the one motor/reel design concept that uses a single high inertia capstan drive and a synchronous capstan motor. Multiple motor reel and drive mechanisms have a tendency to damage the tape, because while driving one reel, the tape must pull the motor of the unused or forward drive reel, thus creating an undesirable back-tension.

Some state-of-the-art characteristics of full-size and miniature tape cassettes and cartridge tape drives are presented in Table 7.2.5.2.1-1.

TABLE 7.2.5.2.1-1. STATE OF THE ART IN CASSETTES AND CARTRIDGE TAPE DRIVES

SYSTEM CHARACTERISTICS	FULL-SIZE CASSETTE	MINI-CASSETTE	FULL-SIZE CARTRIDGE	MINI-CARTRIDGE
Unformatted Capacity (Kbytes)	720 to 2,000	64	2,870	772
Number of Tracks	1, 2, or 4	1	1, 2, or 4	1 or 2
Transfer Rate (Kbits/sec)	8 to 96	2.4	16 to 432	24 to 48
Operating Speed (in/sec)	6 to 30	3	10 to 30	10 to 30
Search Speed (in/sec)	24 to 150	NA	10 to 240	10 to 90
Start/Stop Time (msec)	20 to 120	70	24 to 35	NA
Average Access Time (sec)	20 to 120	14 to 20	14 to 50	11.2 to 21
Average Drive Size (in.)	4 × 6 × 8	3 × 3 × 1.1	3 × 7 × 10	4 × 5 × 4.5
Average Weight (lb)	5	1	4 to 5	1 to 3.2
Average Voltage Requirements (V)	+12, -12, +5	+5	+18, -18, +5	+12, +5
Average Drive Price (qty) Without Controller	$300 to $700	$260	$850	$250
Media Size (in.)	4 × 2.5 × 0.4	2 × 1.3 × 0.3	4 × 6 × 0.665	2.4 × 3.2 × 0.4
Average Media Price	$8.00	$7.70	$18.00	$14.00
Recording Format	ANSI/ECMA	NA	ANSI/IBM Compatible	NA
Tape Length (ft)	Up to 600	Up to 50	Up to 600	Up to 140
Recording Density (bits/in)	800 to 1,600	800	1,600 to 6,400	800 to 1,699
Soft Error Rate	10^{-7}	10^{-7}	10^{-8}	10^{-6}
System Cost (¢/bit)	1×10^{-2}	4×10^{-2}	4×10^{-3}	5.7×10^{-3}
Media Cost (¢/bit)	2×10^{-4}	2×10^{-3}	8×10^{-6}	2.3×10^{-4}
MTBF (hr)	5,000 to 10,000	5,000	10,000	10,000

7.2.5.2.2 Trends in Cassettes and Cartridges - The major push in cassette/cartridge technology today is to overcome the inherently faster access time offered by random-access floppy disk drives by increasing tape speed or increasing the number of tracks for storing data. The other trend, toward miniaturization, has produced several new transports that use a wafer cassette; that is a very small, thin, continuous loop cartridge. This wafer uses a 0.7-in.-wide 50-ft tape that can store 1.44 million flux changes at 2,400 fci, 1.5 Mbits of modified FM (MFM) data, 0.75 Mbit of biphase, or 0.5 Mbit of ratio recording data.

Technological advances in cassettes and cartridges will expand their uses in microprocessor loading applications, data logging, and anlaysis, storage for microprocessor development systems, and portable program loading devices.

A recent trend in cassette/cartridge digital recording is the use of the reel-to-reel hub drive instead of the widely used capstan/pinch roller technique. The main disadvantage of the capstan drives is the hammering and grinding of the tape oxide coat which reduces tape life. The reel-to-reel design can increase tape life by as much as five times. The reel-to-reel hub drive design uses the hub drive, which was previously only used for rewind, fast-forward, and tape-slack operations, to provide a constant tape speed over the read/write head.

As cassettes and cartridges begin using higher tape speeds, multi-gap heads, bidirectional tape motion, and reduced start/stop times, a conflicting force appears. Increased tape and head life require minimum pad pressures, while bidirectional tape motion and small start/stop time require increased tape forces. The evolving problem will have to be overcome by the development of sophisticated pressure pads and tape guiding techniques.

The advent of low-cost microprocessors is putting pressure on the manufacturers to further reduce the price of cassette and cartridge tape units for use in mircroprocessor-based systems. Even though further increases in bit density tend to lower prices, the constant demand for higher performance has the opposite effect.

7.2.5.2.3 Projected Development in Cassettes and Cartridges - The trends in the previous section indicate that cassette and cartridge technology is headed in the direction of increased storage capacity, greater reliability, and faster access times.

By 1979 there will be an increased use of advanced interface electronics to reduce the complexity of interfacing cassette/cartridge drives with the CPU. These interfaces will use microprocessor and TTL-compatible data input and output lines with strobe and separate lines for control of transport operations. This type interface eliminates the user's concern with such transport functions as start and stop time, leader length, encoding/decoding, and the need for external clocking. The use of self-clocking provides speed tolerant recording that essentially eliminates the effects of flutter, wow, and head misalignment that are normally associated with low-cost tape drives.

Other areas of development for tape cartridges and cassettes during the 1980's will be the use of interdependent track operations, four or more track parallel recording, and improved tape operating speed. The interdependent track operation technique sequentially records unit records across the tracks in a continuous record file that is accessible in less than 1/4th the current access time and is easily overwritable. The use of four-track parallel recording provides two important advantages: high data transfer rate and low search time. Head technology, mechanical wear, and magnetic problems, which currently limit maximum cassette/cartridge search speeds, should ease, allowing these speeds to increase to 60 in/sec by 1980. To precisely control increased motor speeds, the transports will probably use a servo loop that includes solid-state optical tachometers.

7.3 DATA PROCESSING SYSTEM COMMUNICATION ELEMENTS

Data processing system communication elements are discussed here in terms of three types of elements considered to be representative of data processing communications in general. The three elements, illustrated in Figure 7-4, are communication controllers, data multiplexers, and I/O channels. Some typical connections of these devices are illustrated in Figure 7.3-1. Specific topics falling within this outline, such as modems, front-end communications processors, data multiplexers, and demultiplexers, are covered elsewhere in this report (see Subsections 4.1.3, 6.3, 9.1.1.2, and 9.2). Thus this section addresses the aspect of data processing system communication elements not covered in other sections of this report: namely, the impact of specific microelectronic devices (chips) on data processing system communication elements.

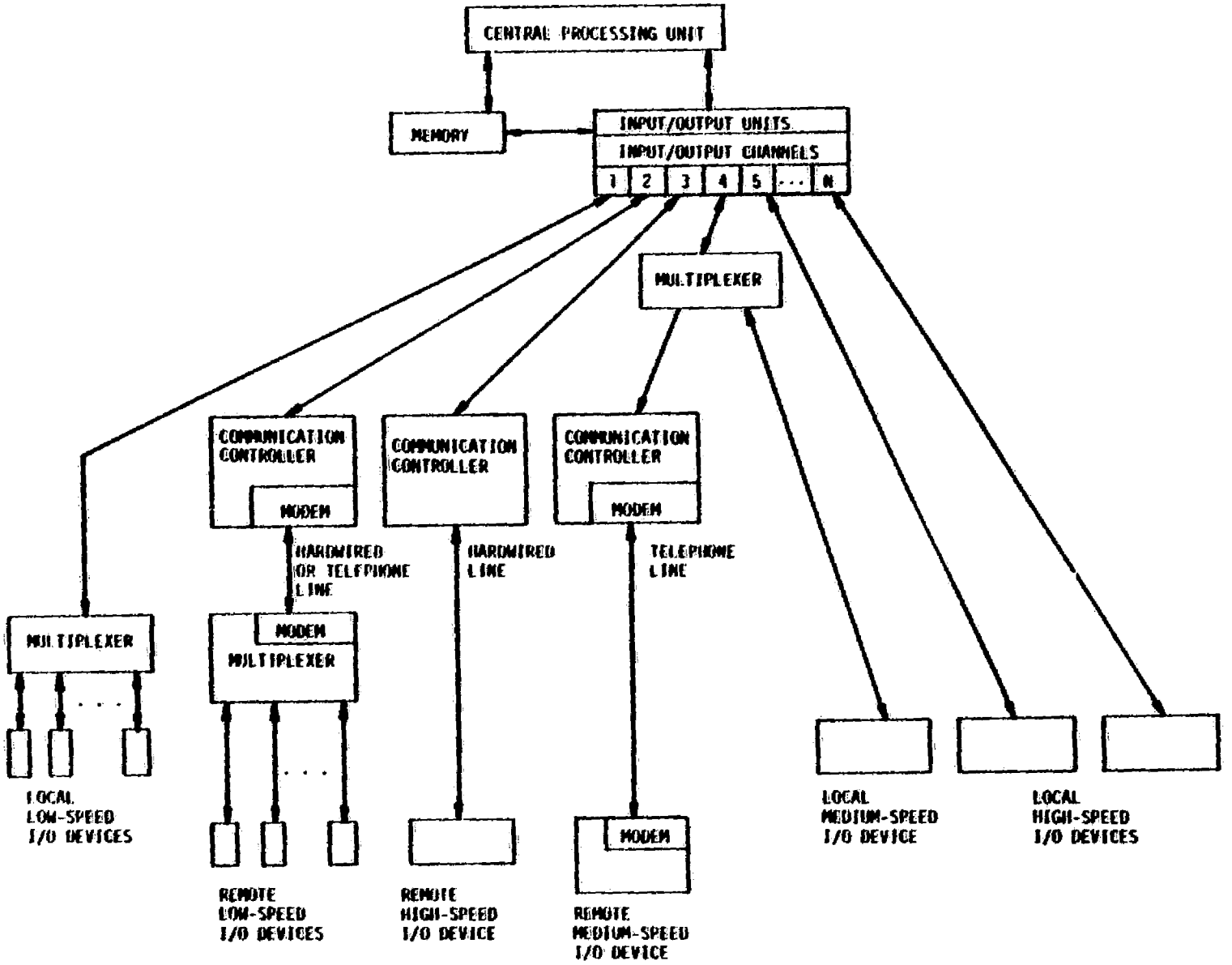


FIGURE 7.3-1. TYPICAL CONNECTIONS OF PROCESSING SYSTEM COMMUNICATION ELEMENTS

7.3.1 State of the Art in Data Processing System Communication Elements

Integrated circuit chips are now available to do much of the work for several types of data processing communication devices. This subsection discusses these available devices (chips and micros) in terms of their capabilities and uses. The logic contained on many of these chips is specifically designed to perform one type of operation (e.g., voltage/frequency translation, serial/parallel data formatting, and multiplex/demultiplex). Microcomputers tailored for communications, on the other hand, are generalized multipurpose devices that may be used to control many communications devices as well as perform a variety of data collection, compression, distribution, format change, etc., tasks.

Communications controllers are those devices that provide timely control of communications between the various elements in a data processing system. Such devices perform such operations as generating and checking error detection/correction codes, buffering data streams, converting data from one format (e.g., voltage levels) to another (e.g., frequency and/or phase of a signal) and back, and converting parallel data to serial (for transmission) and back (for reception). One important microelectronic communications device type currently available, known as a Universal Receiver/Transmitter or Communications Interface, typically provides on a single chip:

- Conversion from parallel data (up to 8 bits) to bit serial data
- Communication in either an asynchronous (start-stop) or synchronous manner (depending on the device selected)
- Generation and checking of parity
- Optional parity suppression
- Generation and detection of start and stop bits (for asynchronous devices)
- Separate clocks for received and transmitted data

- One byte (8 bits) buffering of received and transmitted data (separate buffers for each direction)
- Generation of a signal (interrupt) upon receipt of data
- Recognition of several error conditions, including overrun (receiving new data before the old are removed from the buffer, thus causing data loss) and framing error (receiving a wrong number of bits/word).

Most such devices are programmable to the extent that several features (such as parity and word length) are selectable by writing data into status registers within the device. Typical devices allow transmission at up to 500,000 baud (bits/sec) with external devices or 50,000 baud without them. Another important microelectronic device for communications is the modem chip. This device converts serial data expressed as voltage levels (such as from a Communications Interface) to and from frequency and/or phase of an electrical signal suitable for transmission over voice-grade communications channels. Other typical features include:

- Full or half duplex operation selectable
- Automatic answer
- Automatic disconnect with selectable delay
- Selectable originate or answer mode
- Variable or fixed data rate (depending on device selected).

Typical rates are 0 to 600 baud for variable-rate chips and 2,400 baud for fixed-rate chips. Note that these rates are common for many commercially available modems but are considerably slower than the fastest units available (9,600 baud using a single voice-grade line).

Some communication control devices such as front-end processors and modems are discussed in Section 9.2.

Data multiplexers are those devices that take several streams of data from various sources and combine them into one stream in order to transmit or otherwise process them as a single stream. Demultiplexers, functioning in reverse, break down a single stream into component streams. The term multiplexer (or mux) usually refers to a piece of hardware that

can perform both jobs. Multiplexer (and demultiplexer) chips implemented with ECL logic are available with the capability of handling up to 16 input (output) streams to (from) one output (input) stream at data rates of about 100 Mbits/sec. However, this chip represents only a small part of the logic required in state-of-the-art multiplexers. For complete multiplexers, many other components are required, as buffer storage, control logic, and line drivers, for example. Additionally, several levels of multiplexing may be used to allow a large number of inputs while reducing, to the minimum level, the need for very-high-speed devices used for the logic in the last level of multiplexing.

Data multiplexers (and demultiplexers) are discussed in Subsections 4.1.3, 6.3, and 9.1.1.2.

I/O channels are those devices that provide for data transfer between a processor/computer element and its various input/output devices. They are also known as I/O controllers. I/O channels typically carry one word or byte at a time and include communication lines for parity and/or other error checking/correcting codes. Single-chip devices (Peripheral Interfaces) are now available that provide some of the parts for two I/O channels. They typically provide:

- Two 8-bit data paths, one to each of two devices, and one common 8-bit data path back to the processor/computer element
- Interrupt input signals (from the I/O devices) and output signals (to the I/O devices and the processor/computer element)
- Control registers that hold information such as direction (input or output) for each I/O pin and interrupt status.

They typically do not provide for parity generation or checking, although one of the eight data lines may carry parity information if the data word to be transmitted is 7 bits or less in width. Another important facility supplied by I/O controllers, Direct Memory Access (DMA), can now be provided by a single chip. Thus two chips (a Peripheral Interface chip and a DMA chip) can provide most of the logic necessary for an I/O

channel (minus any error checking/correcting features) capable of transferring approximately 1 Mbyte/sec. Such a controller would be classified as a relatively "dumb" and inflexible device. To provide "intelligent" I/O, computer manufacturers are now using microcomputers in single- and multiple-chip configurations for controlling the I/O processes. Such an intelligent device can relieve the main processor/computer element of a significant portion of its load with respect to I/O by performing such tasks as data compression, checking, collection, routing etc.

Microcomputers are discussed in Subsection 7.1.1. I/O controllers are discussed briefly in Subsections 7.1.3.1 and 7.1.4.1 as they are used in current large-scale and super-scale computers.

7.3.2 Trends and Projected Developments in Data Processing System Communication Elements

The general trend in data processing system communication elements is toward higher data rates and more error check/correction. The general trend of microelectronic devices is toward higher speed and complexity. The device manufacturers are continuing their pursuit of increasing capability and decreasing cost for microelectronic devices. This effort is making it possible to build faster and "smarter" data processing system communication elements.

Some improvements that can be expected in data processing system communication microelectronic devices through the mid-1980's are:

- Incorporation of parity generating/checking in Peripheral Interface chips by 1980
- Incorporation of more sophisticated error correcting codes (e.g., Hamming codes) in Communications Interface and possibly Pheripheral Interface chips before 1985
- Expansion of data widths in some devices (especially Peripheral Interfaces) to 16 bits by 1980 and further expansion to 32 bits by 1985
- More input (output) channels for data multiplexer (demultiplexer) chips to 32 (but not likely 64 because of the number of I/O pins necessary) by the early 1980's
- Modem chips capable of sending and receiving data at 9,600 baud by the early 1980's
- Higher data rates for all devices.

Trends and projected developments for data processing system communication elements (above the chip level) can be found in Subsections 4.1.3.3, 6.3.2, 6.3.3, 9.2.1.2, and 9.2.2.2.

7.4 FIRMWARE

Firmware is composed of two elements: software that is stored permanently and the hardware in which it is stored. Software is discussed in Sections 8 and 11; the hardware, read-only memory, is discussed here.

Read-only memories (ROMs) have been used for many years as microprogram controllers to provide flexible control of CPUs to implement basic instruction sets and frequently used subroutines. The recent dramatic decrease in the cost of semiconductor read-only memory has greatly expanded the use of ROMs in many areas, including microprogram firmware, systems firmware, and applications firmware. Some other historical applications of read-only memories include character generators, function look-up tables, and code converters.

Microprograms are almost always stored in read-only memories to protect them from overwriting (accidental or malicious) and loss during power outages. One exception is writable control store (WCS), which may be classified as read-mostly memory. Fixed microprogram firmware has been used for many years, essentially since computers were invented. The concept of changeable micro-firmware is also very old, but the expense involved has generally precluded its application except in larger processors. With the use of inexpensive semiconductor read-only memories, the concept of changeable or custom microprogram firmware has gained wide acceptance. Custom micro-firmware allows a computer to be optimized for a particular application or instruction mix, or to be set up to emulate other computers. Changeable micro-firmware allows more flexibility in that the optimization or emulation may be refined or changed as requirements change. Such an application might use read-only memory (including erasable/changeable read-only memory) or read-mostly memory (e.g., writable control store).

Systems firmware is systems programs (operating systems, monitors, compilers, library routines, etc.) implemented in firmware. Microcomputers have used small (about 1-Kbyte) read-only memory stored monitor programs for several years, but little work was done on any wide scale in systems

firmware prior to that. Much study of the feasibility of firmware implementation of at least the most used parts of large-scale computer operating systems is currently underway. Results so far look very promising. Presently, compilers and interpreters (8 Kbytes or more in size) for FORTRAN, BASIC, etc., are available in semiconductor read-only memories for several popular microprocessors. Extension of this idea to large-scale computers is also under study. In the past, system libraries (function calculation programs, I/O routines, etc.) have occasionally been kept in read-only memory. This will occur more often and to a larger degree in the future. Note that the subroutine libraries (especially function calculations) have often been implemented at a microprogram level when the need for speed was the driving factor.

Storage of applications programs (firmware) is another use for read-only memories that has found wide-scale application in microcomputers. Many microcomputers are used in a "one-program-for-life" environment as device controllers, etc. In such a system, storage of the program and constants in a read-write memory is undesirable due to overwriting and power outage problems, so that use of read-only memory becomes an obviously superior concept. Moreover, use of inexpensive semiconductor read-only memories will cause firmware realizations of programs of much shorter life cycle to be economically justified. Typical applications program candidates would include those with fairly wide application and moderate lifetime, such as business programs (accounts receivable, general ledger, etc.) and scientific subroutines (linear regression, curve plotting, matrix manipulations, etc.).

The remainder of this section is devoted to a discussion of state of the art, trends, and projected developments in semiconductor read-only memory.

7.4.1 State of the Art in Semiconductor Read-Only Memory

Semiconductor read-only memories may be differentiated along the lines of technology and programming method. The technology used (bipolar, MOS) makes a difference in the speed, power dissipation, and other parameters of the memory. However, differences in programming methods are more striking. Read-only memories may be programmed at the factory, during manufacturing, using one or more masking processes (ROMs); or in the field (PROMs); and some may even be programmed, erased, and reprogrammed (EPROMs, EEROMs, EAROMs). Table 7.4.1-1 summarizes the state of the art in semiconductor read-only memory.

Mask programmed ROMs may be implemented with either bipolar or MOS technology. ROMs of either technology exhibit higher speeds than the corresponding PROMs and also have the lowest cost per bit (for large numbers of identical parts). ROMs are now available that store up to 64 Kbits, more than PROMs or erasable types. Bipolar ROMs are faster but dissipate more power than MOS ROMs. The main disadvantages of ROMs are their total lack of flexibility, their large cost in small volume (due to the fixed overhead of generating the mask for making the chips), and the long leadtime needed to acquire a chip with a new pattern (due to the time needed to generate and verify the mask, and verify sample parts).

Field programmable read-only memories, or PROMs, may also be implemented with bipolar or MOS technology. PROMs are basically a matrix of diodes (or fuses) or fusable links which the user selectively burns out (blows) or fuses by application of large voltage pulses. Although they have slower access times than corresponding ROMs, they can add much flexibility to a system by allowing the user to make rapid changes (a few minutes or less for the actual programming) in firmware.

Total flexibility may be gained by allowing the end user to program, erase, and reprogram read-only memories. Erasability is currently available only for MOS devices, so that a penalty in maximum speed capability must be paid to gain this flexibility. Such memories are always

TABLE 7.4.1-1. STATE-OF-THE-ART IN SEMICONDUCTOR READ-ONLY MEMORY

CHARACTERISTICS	BIPOLAR ROM	MOS ROM	BIPOLAR PROM	MOS PROM	EPROM	EEROM	FAROM
Chip							
Capacity (Kbits)	8	64	8	16	16	8	8
Size (mil²)	NA	NA	NA	NA	NA	NA	44,720
Maximum Read Time (nsec)	35 to 50	350	70 to 250	450	450	450	600 to 2,000
Minimum Cycle Time (nsec)	35 to 50	350	70 to 250	450	450	450	1,800 to 2,500
Maximum Power Dissipation (μW/bit)	104 typical	22 to 10	104 to 40 active 104 to 40 standby	40 to 30 active 40 to 8 standby	40 to 30 active 40 to 8 standby	136 to 118	55 active 25 standby
Cost (¢/bit)	NA	NA	NA	NA	NA	NA	0.3
MTBF	Very High	Very High	Medium	Medium	High	NA	High (medium to low for large number of writes)
System							
Capacity (Mbits)	Up to 64	Up to 256	Up to 64	Up to 128	Up to 128	NA	Up to 18
Physical Volume	NA	NA	NA	NA	NA	NA	NA
Maximum Read Time (nsec)	65 to 90	390	100 to 290	490	490	490	640 to 2,400
Minimum Cycle Time (nsec)	65 to 90	390	100 to 290	490	490	490	2,200 to 3,000
Maximum Power Dissipation (μW/bit)	116 typical	34 to 21	116 to 50 active 116 to 50 standby	54 to 34 active 54 to 13 standby	54 to 34 active 54 to 13 standby	148 to 130	67 active 35 standby
Cost (¢/bit)	NA	NA	NA	NA	NA	NA	NA
MTBF	NA	NA	NA	NA	NA	NA	NA

ORIGINAL QUALITY OF THE PAGE IS POOR

field programmable and are erased in a variety of ways. Erasable programmable read-only memories (EPROMs) are erased by exposure to an intense ultraviolet light source. The erasure is accomplished in a period of 10 or more minutes, and the entire contents of the memory chip are erased at once. Electrically erasable read-only memories (EEROMs) are similar except that they employ electrical signals to erase the entire memory and erase quickly compared with EPROMs. EPROMs and EEROMs use a floating-gate avalanche MOS process. Electrically alterable read-only memories (EAROMs) differ from EEROMs in that they are alterable (reprogrammable) on a word-by-word or block-by-block basis, thus eliminating the need to erase and rewrite all of a chip's contents to make small changes. EAROMs, however, are the slowest of available semiconductor read-only memory types. EAROMs are built using a metal-nitride-oxide MOS process called MNOS. For further details on MNOS, see Subsection 7.2.1.2.

7.4.2 Trends and Projected Developments in Semiconductor Read-Only Memory

Use of read-only memory is expanding, and new, larger, and faster ROMs, PROMs, etc., are constantly being introduced in a highly competitive fashion. These trends should continue through the mid-1980's. Chip density should continue to increase for all types of read-only memories, with bipolar types improving less than MOS types because of the greater maturity of bipolar technology. Access times, cost, and power dissipation per bit should continue to decrease for all types, with MOS again taking the lead. Access times may be expected to provide the most dramatic advances as they fall to the levels of the currently faster RAMs using corresponding technologies. In fact, it is probable that read-only memories will be built speed-optimized rather than chip size, package size, or power optimized. To this end, nonmultiplexed addressing schemes will continue to be used for read-only memories so that access will not have to wait for the presentation of two or more sets of time-multiplexed address bits to the address pins, but may proceed on presentation of the entire address at once.

Reliability will also improve, especially in PROM devices. The future of EAROMs is very uncertain because of questions about fatigue problems associated with the MNOS technology used to fabricate them. Table 7.4.2-1 presents the projected technology in semiconductor read-only memory for the 1980 to 1985 timeframe. In addition to the technologies presented in the table, it is very likely that GaAs ROMs with even faster access times will be available within this timeframe.

TABLE 7.4.2-1. PROJECTED DEVELOPMENTS IN SEMICONDUCTOR READ-ONLY MEMORY, 1980-1985 TIMEFRAME

CHARACTERISTICS	BIPOLAR ROM	MOS ROM	MOS PROM	EPROM
Chip				
Capacity (Kbits)	64	1,024	256	256
Size (mil^2)	NA	NA	NA	NA
Maximum Read Time (nsec)	5 to 35	25 to 100	35 to 100	35 to 100
Minimum Cycle Time (nsec)	15 to 50	55 to 130	60 to 145	66 to 145
Typical Power Dissipation (μW/bit)	75	10	20	20
Cost (¢/bit)	NA	NA	NA	NA
MTBF (hr)	10^8 to 10^9	10^8 to 10^9	10^8	10^8
System				
Capacity (Gbits)	2	2	2	2
Physical Volume	NA	NA	NA	NA
Maximum Read Time (nsec)	15 to 50	35 to 115	45 to 120	45 to 120
Minimum Cycle Time (nsec)	20 to 60	60 to 180	70 to 150	70 to 150
Typical Power Dissipation (μW/bit)	85	15	25	25
Cost (¢/bit)	NA	NA	NA	NA
MTBF (hr)	10^5	10^5	10^4 to 10^5	10^4 to 10^5

REFERENCES - SECTION 7

7-1 H. S. Stone et al., Introduction to Computer Architecture, Science Research Associates, Inc., pp. 143-150, 1975

7-2 D. A. Hodges, "Trends in Computer Hardware Technology", Computer Design, pp. 77-81, February 1976

7-3 F. G. Withington, "Beyond 1984: A Technology Forecast", Datamation, pp. 54-65, January 1975

7-4 F. Vaughan, "Continuing Increases in Price/Performance Seen for Large User", Computerworld, April 11, 1977

7-5 "Air Force Command and Control Information Processing in the 1980's: Trends in Hardware Technology", The Rand Corporation, Report R-1011-PR, Santa Monica, CA, October 1972

7-6 C. P. Lecht, "The Waves of Change", Computerworld, pp. 9-11, April 18, 1977

7-7 H. Falk, "Computer Systems: Hardware/Software", IEEE Spectrum, pp. 38-43, January 1975

7-8 A. Lynch, "Distributed Processing Solves Mainframe Problems", Data Communications, pp. 17-22, November/December 1976

7-9 C. Adams, "Over the Horizon", A Report on the June 1975 Computer Elements Technical Committee Workshop, Computer, pp. 8-11, February 1976

7-10 T. A. Dolotta et al., Data Processing in 1980-1985, pp. 67-76, March 1976

7-11 D. House, "CCD Versus RAM for Bulk Storage Applications", IEEE 1976 Compcon, pp. 58-61

7-12 R. Allan, "Semiconductor Memories", IEEE Spectrum, pp. 40-45, August 1975

7-13 W. S. Boyle and G. E. Smith, "Charge Coupled Semiconductor Devices", Bell System Technical Journal, Briefs, 49, p. 587, April 1970

7-14 A. F. Tasch, Jr., H. S. Fu, T. C. Holloway, and R. C. Frye, "Charge Capacity Analysis of the Charge-Coupled RAM Cell", IEEE Journal of Solid-State Circuits, Vol. SC-11, pp. 575-584, October 1976

7-15 A. F. Tasch, Jr., R. C. Frye, H. S. Fu, "The Charge-Coupled RAM Cell Concept", IEEE Journal of Solid-State Circuits, Vol. SC-11, pp. 58-63, February 1976

7-16 J. McDermott, "Charge-Coupled Approach Leads to New LSI Digital Logic Devices", Electronic Design, pp. 20-22, February 1, 1977

7-17 W. C. Hughes et al., "BEAMOS - A New Electronic Digital Memory", National Computer Conference, pp. 541-548, 1975

7-18 D. E. Speliotis et al., "Electron Beam Memories", Electro 76 Professional Program, Section 33-3, pp. 1-12

BIBLIOGRAPHY - SECTION 7

J. Koudela, Jr., "The Past, Present, and Future of Minicomputers: A Scenario", Proceedings of the IEEE, Vol. 61, No. 11, pp. 1526-1534, November 1973

R. Allan, "Components: Microprocessors Galore", IEEE Spectrum, pp. 50-56, January 1976

F. O. Arntz et al., "Electron Beam - Accessed Data Storage in MOS Capacitors", IEEE International Solid-State Circuits Conference, pp. 176-177, 1976

C. G. Bell et al., "Effect of Technology on Near Term Computer Structures", Computer, pp. 29-37, March/April 1972

D. M. Bowers, "Minicomputers Part 1 - History and Market Outlook", Mini-Micro Systems, pp. 40-42, May 1976

D. M. Bowers, "Minicomputers and Microcomputers Part 3 - Minicomputers", Mini-Micro Systems, pp. 32-40, August 1976

J. W. Bremer, "Hardware Technology in the Year 2001", Computer, pp. 31-36, December 1976

D. Bursky, "What's Ahead in 1977", Electronic Design, pp. 34-36, January 4, 1977

D. Christiansen, "Technology '76'", IEEE Spectrum, pp. 42-43, January 1976

H. Falk, "Computers: Poised for Progress", IEEE Spectrum, pp. 44-49, January 1976

H. Falk, "Reaching for a Gigaflop", IEEE Spectrum, pp. 65-69, October 1976

H. Falk, "Computers: All Pervasive", IEEE Spectrum, pp. 38-43, January 1977

G. C. Feth, "Progress in Memory and Storage Technology", Electro 76 Professional Program, Section 33-4, pp. 1-11

R. A. Frank, "All Signs Pointing to a New Generation", Computerworld, pp. 1-2, April 18, 1977

R. A. Henle et al., "The Application of Transistor Technology to Computers", IEEE Transactions on Computers, pp. 1289-1303, December 1976

D. A. Hodges, "A Review and Projection of Semiconductor Components for Digital Storage", Proceedings of the IEEE, Vol. 63, No. 8, pp. 1136-1146, August 1975

W. C. Hughes et al., "Electron Beam Accessed Mass Memory", Proceedings of the IEEE, Vol. 63, No. 8, pp. 1234-1240, August 1975

D. P. Jasper, "Position Paper on Usability for Data Services (Why Not a Mini-Computer in Every Cupboard)", IEEE 1975 Compcon, pp. 65-69

T. W. Keller, Cray-1 Evaluation Final Report, LA-6456-MS, Los Alamos Scientific Laboratory, December 1976

J. Kelly, "The Development of an Experimental Electron-Beam-Addressed Memory Module", Computer, pp. 32-42, February 1975

P. J. Martin, "The Effect of Hardware on Systems Software 1975-1985", IEEE 1975 Compcon, pp. 337-339

R. R. Martin and H. D. Frankel, "Forecast of Computer Memory Technology", IEEE 1974 Compcon, pp. 287-290

R. R. Martin and H. D. Frankel, "Electronic Disks in the 1980s", Computer, pp. 24-30, February 1975

W. Myers, "Key Developments in Computer Technology: A Survey", Computer, pp. 48-77, November 1976

R. A. Pedersen, "Integrated Injection Logic: A Bipolar LSI Technique", Computer, pp. 24-29, February 1976

A. V. Pohm, "Electronic Replacements for Head-per-Track Drums or Disks", Computer, pp. 16-19, March 1976

H. Schmid, "Monolithic Processors", Computer Design, pp. 87-95, October 1974

P. Schreier and J. Conway, "Design News", EDN, pp. 29-32, April 20, 1977

D. E. Speliotis, "Bridging the Memory Access Gap", National Computer Conference, pp. 501-508, 1975

R. Sugarman, "Memory Report", Electronic Engineering Times, pp. 26-33, April 4, 1977

D. J. Theis, "Vector Supercomputers", Computer, pp. 52-61, April 1974

D. J. Theis, "The Midicomputer", Datamation, pp. 73-82, February 1977

E. A. Torrero, "Solid-State Devices", IEEE Spectrum, pp. 48-54, January 1977

P. W. J. Verhofstadt, "Technology Constraints, Present and Future", Computer, pp. 32-35, January 1976

R. J. Whittier, "Semiconductor Memories: The Impact and Momentum of Current Technology", Electro 76 Professional Program, Section 33-1, pp. 1-4

J. F. Wirsching, "Computer of the 1980's - Is It a Network of Microcomputers?", IEEE 1975 Compcon, pp. 23-26

K. Yoshida et al., "A 16-Bit LSI Minicomputer", IEEE Journal of Solid-State Circuits", pp. 696-701, October 1976

"What the Experts Say", Data Processing, pp. 45-47, November 1976

"Data Topics", Electronic News, December 13, 1976

"Compcon Abstracted", Mini-Micro Systems, pp. 44-51, October 1976

Datapro 70, ©1976 Datapro Research Corporation

"Semi-Vendor Tool-up for Next Generation of RAMs", Electronic Engineering Times, pp. 10, February 7, 1977

"Super-Super Computer", Datamation, pp. 150-152, March 1977

"Electro 77 Another Show - Another Opening", Electronic Design, pp. 54-64, April 12, 1977

"Gigabit CRT Seeks Memory Niche", Electronic Engineering Times, pp. 30-33, April 4, 1977

A. J. Kurtzig et al., "Looking into the Future", Computer, pp. 14-15, March 1976

W. H. Herndon et al., "A Statre 4096-bit Bipolar Random Access Memory", IEEE Journal of Solid-State Circuits, pp. 524-527, October 1977

"Static Induction-Transistor Cell May up RAM Performance Tenfold", Electronics, p. 64, September 15, 1977

P. M. Quinn et al., "A 16K × 1 I^3L Dynamic RAM", 1978 IEEE International Solid-State Circuits Conference, pp. 154-155

R. F. Broom et al., "Model for a 15 ns 16K RAM with Josephson Junctions", 1978 IEEE International Solid-State Circuits Conference, pp. 60-61

D. J. Theis, "An Overview of Memory Technologies", Datamation, pp. 113-131, January 1978

L. Altman, "New MOS Processes Set Speed, Density Records", Electronics, pp. 92-94, October 27, 1977

"Japanese 4-k RAM Challenges U.S. Devices", Electronics, pp. 39-40, September 15, 1977

F. B. Jenné, "Grooves Add New Dimension to V-MOS Structure and Performance", Electronics, pp. 100-106, August 18, 1977

L. Altman, "65-k RAMs Won't Slight Performance", Electronics, pp. 80-82, February 16, 1978

T. Sugano et al., "Solid State: Synonym for Progress", IEEE Spectrum, pp. 51-56, September 1977

R. Allan, "Semiconductors: Toeing the (Microfine) Line", IEEE Spectrum, pp. 34-40, December 1977

K. Shimotori et al., "Fully Ion Implanted 4096-Bit High Speed DSA MOS RAM", 1977 IEEE International Solid-State Circuits Conference, pp. 76-77

8. DATA BASE SYSTEMS ELEMENTS (GROUND)

Data base systems elements include all the Level 1 elements presented in Figure 8-1. This section discusses only two of the elements identified in the referenced figure: i.e., Data Base Management System (DBMS) and Data Storage Elements. Processing is discussed in Section 7; telecommunication elements are presented in Section 9; and software technology is included in Section 11.

Figure 8-2 presents a detailed breakdown of DBMS elements in terms of features and functions that are generally used to characterize a DBMS. Subsequent paragraphs are organized along the lines of the breakout presented in this figure.

DBMS represents perhaps the most complex data processing technology in the industry today. This is attributable to the several disciplines that are crossed to develop, assess, or even understand DBMSs as they exist today. As a result, two constraints or ground rules were established to effect a meaningful technology assessment within the limitations of this contract effort.

The first constraint is related to the scope of the subject matter. One alternative is to cover all aspects of DBMSs but restrict the level of detail and thereby produce little quantitative results. Another alternative is to concentrate the investigative effort on selected areas considered to be key DBMS issues. This latter approach was taken to assess the technology from the present time through 1985. Although a reasonably comprehensive outline of DBMS issues is contained in this section, not all of these issues were selected for in-depth investigation and assessment.

Second, to provide consistency and commonality to this effort and to enhance communication with experts and authorities in the field, DBMSs were limited to those that are, or will be, commercially available and general-purpose. In the evaluation and assessment of DBMS functions,

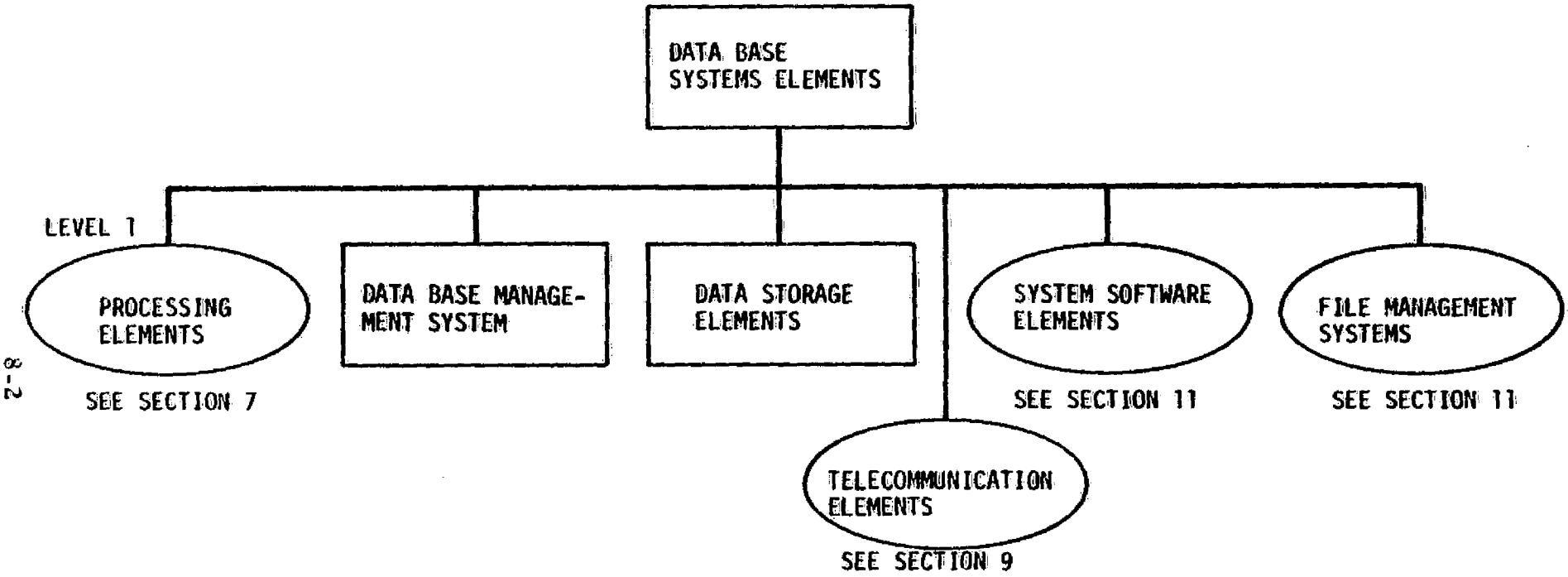


FIGURE 8-1. DATA BASE SYSTEMS ELEMENTS

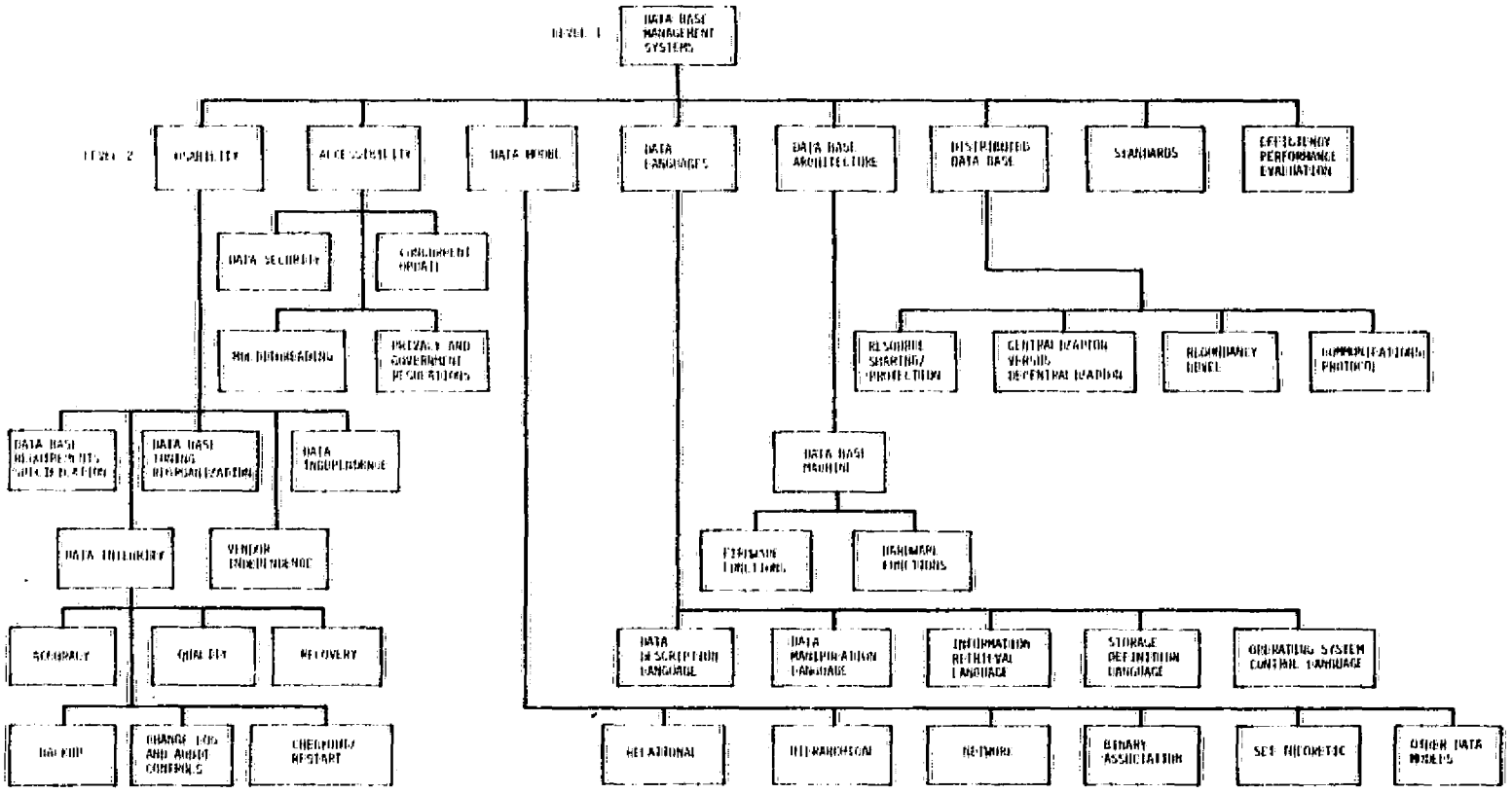


FIGURE 8-2. DATA BASE MANAGEMENT SYSTEM FEATURES AND FUNCTIONS

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

the results are taken to represent those functions and capabilities that are, or will be, common to the majority of commercially available, general-purpose DBMSs.

The data storage elements discussed in Subsection 8.9 include the large-capacity magnetic and optical mass storage elements that are unique to very large data base storage applications. Other types of mass storage devices such as disk, which provide less storage capacity and are thus also used for other data storage applications, are discussed in Section 7.

8.1 USABILITY

The term usability is intended to represent a class of functions that enable the Data Base Administrator's (DBA) staff to effectively maintain, alter, and access the data base to ensure availability of meaningful data to end users. As such, usability entails several aspects which are considered in the following paragraphs.

8.1.1 Data Base Tuning/Reorganization

The purpose of data tuning and/or reorganization is to provide cost-effective performance of both the DBMS software and the set of applications programs that utilize it.

8.1.1.1 State of the Art in Data Base Tuning/Reorganization - Most DBMSs today have a Data Description Language that provides some capacity for logically restructuring the data base. At present, however, the DBMS maintains statistics on use, but the DBA must use the statistics in the reorganization decision-making process.

8.1.1.2 Trends in Data Base Tuning/Reorganization - The whole problem area of which statistics to gather and how often to gather them is also an integral part of the question of performance evaluation (Section 8.8). Much work is under way to identify those operating parameters affecting DBMS performance (Ref. 8.1), but there is little evidence to indicate that sufficient emphasis has been placed on incorporating advanced techniques in commercially available DBMSs. Although there are trends toward continued development in this area, there appears to be insufficient market pressure to warrant a rapid development pace.

8.1.1.3 Projected Developments in Data Base Tuning/Reorganization - In terms of capabilities, it is envisioned that by 1985 (and no sooner) the DBMS and the DBA will jointly use the DBMS in the sense that the DBA can define a utilization statistics data base in much the same manner as any other data base might be defined. The DBA, as user, would then access and analyze these data, which are automatically maintained by the DBMS. Because of the obvious magnitude of resources required, this capability will be provided only on the largest installations for some time after initial development.

8.1.2 Data Independence

Data independence is considered one of the key DBMS issues today. Independence means the ability to make a change to the data base (whether logical or physical) without significantly affecting the programs that access it. The degree to which the programs must be changed represents the degree of data dependence (i.e., more change implies greater data dependence).

For this effort, data independence was classified into two general categories. The first category is termed relative independence. A collection of application programs is said to be relatively data-independent with respect to a change if the only required changes are in those programs that directly access the changed view or structure of the data base. The second category is termed absolute independence. A collection of applications programs is said to be absolutely data-independent with respect to a change if it is not necessary to change or recompile any of the existing applications programs to enable their execution. To do this, in the case of data item deletion, for example, it is necessary for the DBMS to recognize an undefined data item and use some type of algorithm to substitute a default value upon accessing the data base to provide the application program an executable environment. Obviously, the long-term ramifications of such a capability imply a level of sophistication not readily achievable in the immediate future.

8.1.2.1 State of the Art in Data Independence - Tables 8.1.2.1-1 and 8.1.2.1-2 provide a summary of the present capabilities with respect to data independence. These tables exhibit different types of data base changes and the corresponding timeframe in which the designated type of independence will exist. Table 8.1.2.1-1 shows that there is a high degree of relative data independence in the present state of the art. Table 8.1.2.1-2, on the other hand, shows that there is need for significant development to achieve the same degree of absolute data independence.

TABLE 8.1.2.1-1. RELATIVE DATA INDEPENDENCE

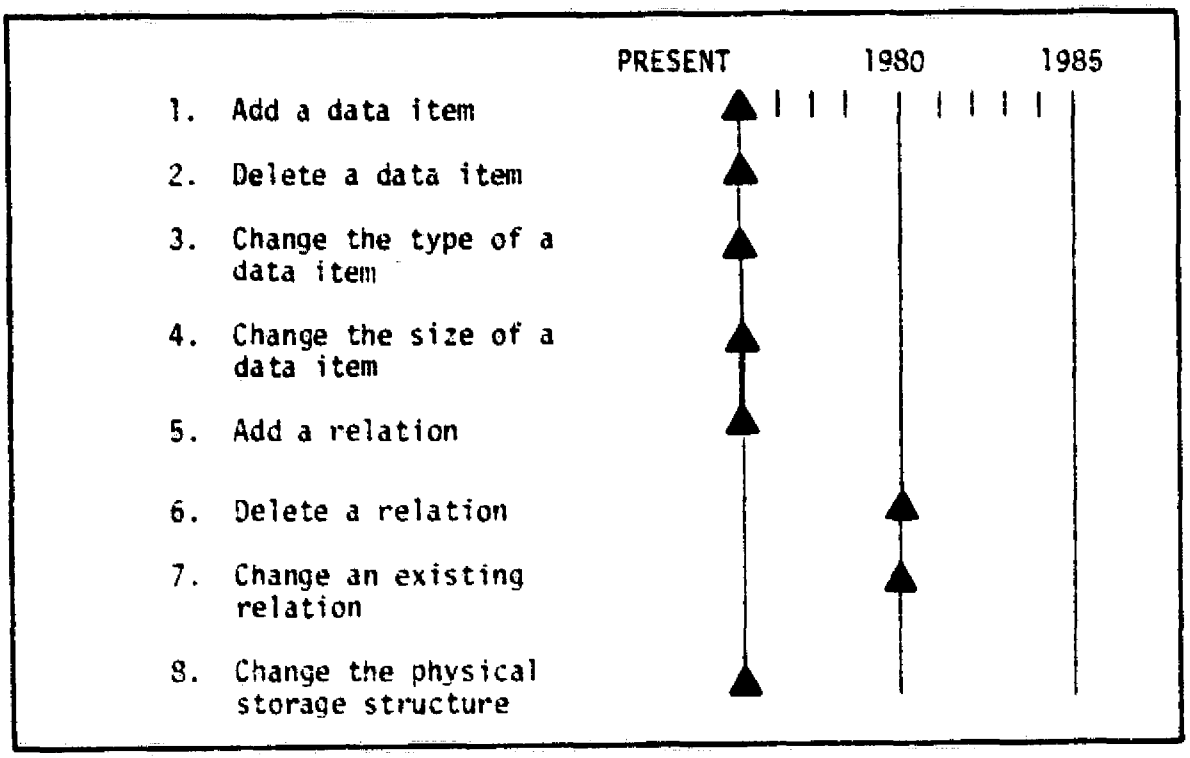
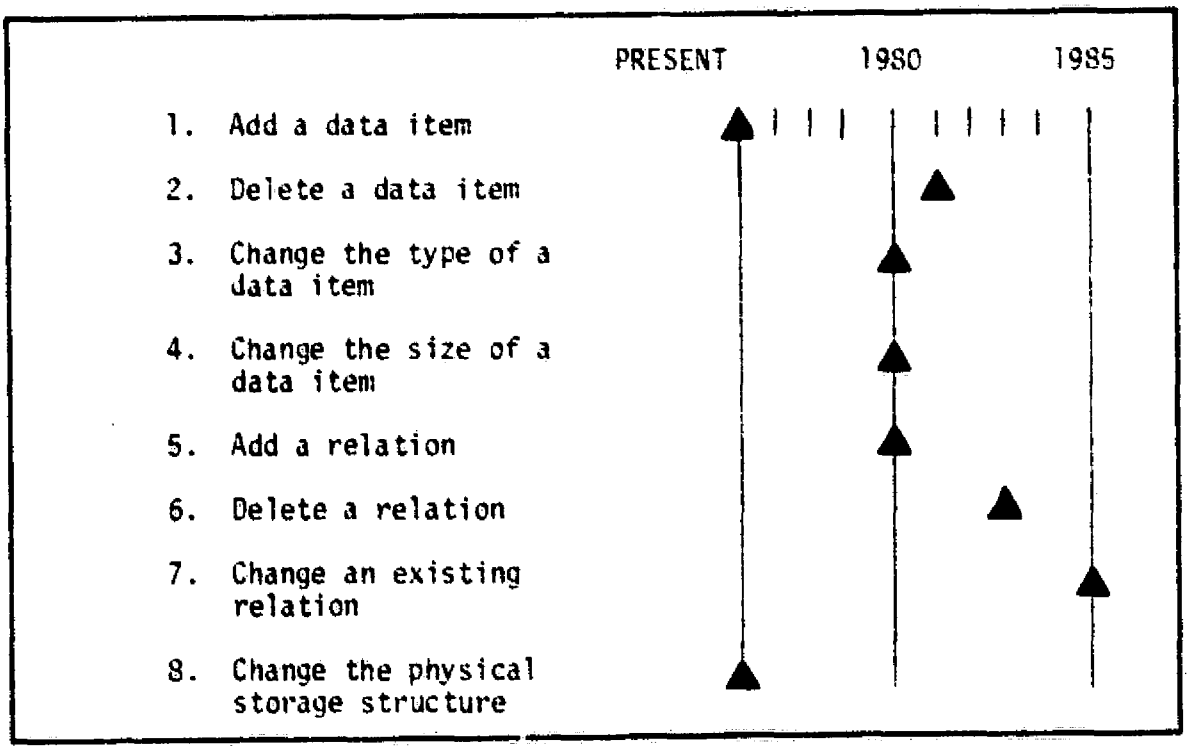


TABLE 8.1.2.1-2. ABSOLUTE DATA INDEPENDENCE



8.1.2.2 Trends in Data Independence - By comparing Tables 8.1.2.1-1 and 8.1.2.1-2, it can be seen that current development trends are much stronger in favor of changes affecting data items as opposed to changes affecting logical data relations. This applies to both relative and absolute data independence. Also, due to the less complex technological considerations, the trend toward comprehensive relative data independence is evidenced by Table 8.1.2.1-1.

8.1.2.3 Projected Developments in Data Independence - It is expected that by 1982 existing application programs will essentially be "immune" to data item changes, as well as to adding a relation. Absolute data independence with respect to data relations is projected to be feasible by 1985. However, as a result of the more extensive ramifications associated with altering relations, it is uncertain at present as to what degree such relational changes can be performed and still accomplish meaningful processing. (An extreme example is the deletion of all relations, in which case any processing by existing applications programs would be meaningless unless for testing/checkout purposes or unless processing is performed on unconnected data.)

8.2 ACCESSIBILITY

Accessibility includes both manual and automated procedures affecting an individual's or organization's ability to look at or alter data. This further means that procedures may vary depending on the mode of access (e.g., batch versus on-line), as well as on the nature of the various data items. The following paragraphs discuss the various aspects of accessibility in the context of DBMS.

8.2.1 Data Security

Data security is considered to be the protection of data against unauthorized dissemination or tampering, whether it be accidental or intentional. Typically, attention is focused on the relationship between the data and users of this data (both authorized and unauthorized).

8.2.1.1 State of the Art in Data Security - One technique employed to discourage unauthorized comprehension of data in a data base is the use of data enciphering/deciphering. Presently, this is not a widespread feature of DBMS (e.g., only one or two DBMSs offer such a feature).

Another technique employed by DBMS to accomplish some degree of data security is the use of various passwords, algorithms, and identifications to lock DBMS resources. In this context, a resource is both the data and the operation performed on that data. Table 8.2.1.1-1 depicts the timeframe when it is predicted that the various DBMS resource levels will be protected from unauthorized access. As shown in this table, the data base and file levels presently have reasonably comprehensive lockout capability. To date, the processing and management of lockout features resides in the DBMS software.

8.2.1.2 Trends in Data Security - The trend in data security is two-fold: 1) greater protection at lower resource levels and 2) the use of hardware to enhance prevention of unauthorized access to data. The use of technological advances in hardware will make the use of enciphering/decphering, multi-level resource lockout, etc., not only feasible but economically practical as well.

8.2.1.3 Projected Developments in Data Security - As shown in Table 8.2.1.1-1, protection at the record level will be commonplace by 1980. At the lower data levels of data aggregate and data item, development will continue so that by 1984 these data levels may also be locked. Almost all of the resource locking capabilities will be in the DBMS software for the immediate future, since widespread use of hardware firmware to replace current and anticipated DBMS software functions is not envisioned prior to the 1983-1985 timeframe.

With rapid hardware developments and the Federal Government's recent development of an enciphering/deciphering standard, it is believed that by 1985 such a feature will be widespread, but as an extra-purchase option.

TABLE 8.2.1.1-1. PROJECTION OF DBMS RESOURCE PROTECTION

OPERATION / DATA LEVEL	"LOOK AT" (READ ONLY)	MODIFY	DELETE	ADD
Data Base	Present	Present	Present	Present
File	Present	Present	Present	Present
Record	1980	1980	1980	1980
Data Aggregate	1983	1983	1983	1983
Data Item	1984	1984	1984	1984

8.2.2 Concurrent Update

A classic problem, which must be solved for any multiprogramming or multiuser environment, is the protection of data against simultaneous modification by two or more users or applications programs.

8.2.2.1 State of the Art in Concurrent Update - Presently, DBMS provides concurrent update protection at the data base and file levels. This is provided in the DBMS software, which is typically reentrant in nature.

8.2.2.2 Trends in Concurrent Update - The trend in this area is to develop techniques that enable protection of the data at the lowest level possible. This frees the remaining data for use by others. A major consideration in this assessment is when and at what level concurrent update protection is expected to be economically practical.

8.2.2.3 Projected Developments in Concurrent Update - By 1979, this protection will be a common feature at the record level. However, concurrent update protection at the data aggregate and data item levels is not expected to be available until 1985.

8.2.3 Privacy and Government Regulations

The Privacy Act of 1974 and other privacy legislation currently pending have a definite impact on the development and trends in DBMS. Since the area of privacy and Government regulations is not technological in nature, it is not included in this effort. However, it is recommended that this area be included in future assessment efforts.

8.3 DATA MODEL

A "data model" is the logical structure with which the user interrelates data items of interest to represent the "real world". DBMSs are typically characterized by the type of data model that they support. For assessing DBMS technology for the future, a DBMS is labeled "network" if the network data model is the primary model around which the DBMS was conceived and designed. For example, INQUIRE might be considered today to be a relational DBMS in light of some of its capabilities. However, literature to date indicates that INQUIRE's present capabilities as a DBMS have evolved from what was originally an information retrieval system. Thus INQUIRE would not be counted among the relational DBMSs projected to be developed in the future.

8.3.1 State of the Art in Data Models

Table 8.3.1-1 gives the distribution of data models among commercially available DBMSs; Table 8.3.1-2 gives the expected distribution among DBMS users. Although there are potentially many structural models into which data may logically fit, these tables show that the three most popular at present are hierarchical, network, and relational.

8.3.2 Trends in Data Models

Tables 8.3.1-1 and 8.3.1-2 illustrate a gradual trend away from hierarchical and other models in favor of the relational and network models.

TABLE 8.3.1-1. DATA MODEL DISTRIBUTION AMONG DBMSs

DATA MODEL	PRESENT	1980	1985
Relational	2%	18%	25%
Network	36%	39%	42%
Hierarchical	58%	40%	28%
Binary Association	2%	1%	1%
Set-Theoretic	1%	2%	3%
Other	1%	0%	1%
	100%	100%	100%

TABLE 8.3.1-2. DATA MODEL DISTRIBUTION AMONG DBMS USERS

DATA MODEL	PRESENT	1980	1985
Relational	2%	14%	21%
Network	34%	36%	39%
Hierarchical	60%	47%	36%
Binary Association	4%	2%	2%
Set-Theoretic	0%	1%	2%
Other	0%	0%	0%
	100%	100%	100%

8.3.3 Projected Developments in Data Models

Although Tables 8.3.1-1 and 8.3.1-2 indicate increasing development and use of the relational model, network is expected to be the primary data model available and in use through 1985. The information in these tables is based on an assessment of responses received from authorities in the field who were asked to predict the future data model distributions.

Some additional notes are necessary to further qualify these projections. First, 0% does not mean "not at all". Rather, 0% is intended to mean either that the number making up the percentage is so small as to be negligible (e.g., less than 1%) or that it reflects special DBMSs or users not considered in this assignment. Second, there is currently a distinct trend towards an increasing number of DBMSs designed around the network model (e.g., CODASYL). Also not shown in the tables is the expectation among experts that DBMSs of the future will support multiple data models. Exactly when such a DBMS might appear on the market is not clear, but prototypes are anticipated to be in development by 1985. Third, the trend away from hierarchical models to network models raises the question of translating a data base from one model to another. To facilitate this process, at least one high-level translation language is expected to exist by 1983. It will be specialized (e.g., to translate IMS to IDMS on IBM), and such translation languages when they exist will continue to be specialized through 1985 and perhaps beyond.

8.4 DATA LANGUAGES

The term data languages refers to the languages that are designed to: 1) define a data base physically and/or logically; 2) allow the data base to be modified, updated, and reorganized; or 3) allow various classes of users to retrieve data and generate reports. The physical organization of the data base is described by physical storage structures such as files and records. A storage definition language is used to define this physical structure. It represents the view of the systems programmer and systems designer, who are concerned with performance, positioning of data on devices, and methods of indexing data. The overall logical data base description is referred to as a schema. The logical description of a portion of the data base, which may be one user's view of the data base, is called a subschema. A data description language is used to define schemata and subschemata. The interface between a user and the data base is a data manipulation facility. With it the user may select data and cause the selected data to be transferred physically between a data base and an application program (or vice versa). The data manipulation facility may support interrogation, query, data manipulation, and report generation languages. A good introduction to the role played by different data languages may be found in Reference 8-3.

8.4.1 State of the Art in Data Languages

Current DBMSs use widely differing means of specifying the physical data structure of their data base. In most existing DBMSs, a separate storage definition language does not exist. Many of the functions of the storage description language are performed by the data definition language, while others are handled by utilities supplied with the DBMS. Proponents of a separate storage definition language point out its utility in providing physical data independence (i.e., the physical layout of data could be changed without affecting the overall logical structure of the data or the applications programs).

Under the influence of the proposed CODASYL standards (Ref. 8-4), a number of software products with data description language facilities are coming onto the market. Most currently available data definition languages follow at least the spirit of the CODASYL standard [e.g., the Control Data Corporation DDL (version 1.0), the Burroughs Data Management System II, and the Sperry-UNIVAC data management system]. One important exception is IBM's Data Language/1 (DL/1). The tree is the basic construct of DL/1, while the set is the basic construct of the CODASYL data description language (Ref. 8-5).

The user of a data base may have access to the data base, either on-line or off-line, through several different languages. A language, such as COBOL or FORTRAN, which has been augmented to include data manipulation features may be used. Alternatively, a self-contained data language (designed specifically for use with the data base) may be used to obtain and manipulate data. A fast-growing subset of the self-contained data languages is the set of English-like query languages. With these languages a user can formulate (usually on-line in an interactive mode) common requests in a simple yet concise manner. A few systems support effective data base interrogation techniques through man-machine dialogue.

8.4.2 Trends in Data Languages

Except for the trend toward providing separate storage definition and data description languages, trends in these two language areas are difficult to distinguish. There does seem to be a trend toward including more manipulative capability in data description languages. It has been suggested in the literature that a complete query language should be included as a proper subset of a data definition language in order to achieve multiple views of the data base.

Several currently available data base management systems provide both host language (in some cases, several host languages) and self-contained facilities for accessing the same data. Since many organizations have need for both type facilities, more systems that support both can be expected to appear. In fact, there is a trend not only to support both type facilities but to support several different languages aimed at different groups of users. For example, a data base may be accessible to 1) an untrained casual user through a query language supporting computer-initiated dialogue, 2) a trained office worker through a programming-like language, and 3) an applications programmer through a host language.

The languages query-by-example, with its fixed column input form, and FORAL LP, with its light pen display screen input, have provided innovations that attempt to make the forming of access requests easier for the user (Ref. 8-6). This is a continuation of the trend to simplify the user interface to the data base.

8.4.3 Projected Developments in Data Languages

In the future, most DBMSs will provide separate, though related, storage definition and data description languages. Storage definition languages will probably remain unique to each vendor, although proposed standards will have some influence.

By 1985, query languages will be in use that have the capability to communicate with casual (untrained) users in what will appear to be natural language. These query languages will support complex linguistic constructions as opposed to the simple constructions supported by prototype systems today. Through dialogue with the user, the system will be able to determine the user's information needs. Although the user trend is toward higher-level languages, for years to come it will continue to be necessary to program the large, repetitive systems of Government and industry efficiently by using host languages such as COBOL and FORTRAN.

8.5 DATA BASE ARCHITECTURE (HARDWARE)

A number of varied advances in the hardware technology will definitely impact the trends in DBMSs of the future. As a result of the increased prominence of DBMSs in the industry and the many computer-consuming functions in DBMSs, experts are anticipating the imminent development of a data base machine.

8.5.1 Data Base Machines

8.5.1.1 State of the Art in Data Base Machines - There are presently computers oriented toward DBMS, such as Datacomputer developed by Computer Corporation of America. This is basically a minicomputer operating as a "back-end" processor to perform data access and organization functions. A similar development is being undertaken by Cullinane. Reference 8-7 provides a summary of the use of dedicated DBMS processors in ongoing work. If these qualify as data machines, then data machines are already available or certainly will be in the very near future. However, since these examples demonstrate new uses for current general-purpose hardware technology rather than a new development in the hardware itself, the "back-end" processor systems being developed today do not qualify as data base machines for this assessment effort. Thus a data base machine does not exist today.

8.5.1.2 Trends in Data Base Machines - In a recent article, Baum and Hsiao (Ref. 8-8) provide an excellent summary of current research and anticipated hardware technology trends which will significantly influence the characteristics of a data base machine. To summarize their view, the data base machine will have electronic mass storage (in conjunction with disk) and multiple, specialized processors and/or memory to perform DBMS functions (e.g., security checks). Other authorities consulted as a part of this effort further indicate that initially the specialized processors will perform conventional distinct functions that can be readily identified in present software. Examples include: searching, sorting, security checking, data base accessing by key, hashing, and some editing of data. After a period of development, the machine described by Baum and Hsiao (Ref. 8-8) exhibiting hierarchical memory, sophisticated data clustering, and a full complement of subsidiary DBMS processors will emerge.

8.5.1.3 Projected Developments in Data Base Machines - Exactly when the first data base machine will exist is not clear because of the vague idea of what characterizes such a machine. However, as qualified in Sections 8.5.1.1 and 8.5.1.2, the first data base machine is not expected to appear on the market before 1982. Also, historical data indicate that only the largest mainframe vendors have the resources to introduce such a potentially radical development and succeed at generating widespread interest.

8.6 DISTRIBUTED DATA BASES

Since there is a diversity of opinion as to what is required for distributed data bases, an attempt was made to define those DBMS features that are considered to be minimal to support distributed data base processing, and then to project when a DBMS with those features will exist.

In assessing distributed data base processing, the following DBMS characteristics were considered to be minimal:

- Concurrent update protection at the record level
- Security protection at the data item level
- Multilevel access authority controls
- Redundancy controls
- Transaction switching/routing
- Global data base dictionary with centralized control
- Distributed recovery capability (plus some degree of centralized recovery).

Other features that might be desirable include: support of multiple data models, reentrant software, multithreading, standard communciations interfaces and protocols, monitoring of data usage, dynamic file allocation, retrieval by content, and modular hardware architecture.

8.6.1 State of the Art in Distributed Data Bases - There is not a commercially available, general-purpose DBMS in existence today that is designed to fulfill the minimal requirements of a distributed data base.

8.6.2 Trends in Distributed Data Bases

Current trends indicate that a variety of efforts will be devoted to developing the individual features described above without particular consideration of integrating all of them in a single DBMS.

8.6.3 Projected Developments in Distributed Data Bases

The results of the survey of experts indicate that a DBMS possessing all of those features considered minimal to support distributed data base processing will not appear before 1985. Isolated systems with partial capabilities will begin to appear, however, in the 1980-1982 timeframe.

8.7 STANDARDS

In the proceedings of the workshop co-sponsored by NBS and ACM (Ref. 8-9), the working panel on standardization reported four prominent areas related to DBMS for which standards are needed: terminology, criteria for evaluation of standards, components (e.g., data description language), and usage.

8.7.1 State of the Art in Standards

To date, the only viable DBMS standards proposal has been that of the CODASYL Data Base Task Group (DBTG). The DBTG report of April 1971 contains the specification of a schema langauge, a COBOL subschema language, and a COBOL data manipulation language oriented toward the network data model.

8.7.2 Trends in Standards

As a general rule, the next 5 years are expected to demonstrate an increasing interest in DBMS standards developments. In addition to the DBTG, there is also an ANSI subcommittee developing similar subschema and data manipulation capabilities for FORTRAN. The status of this committee is not presently known.

8.7.3 Projected Developments in Standards

Many experts expect the proposed standards of the DBTG (or a moderately revised version thereof) to receive ANSI approval by 1980.

8.8 EFFICIENCY/PERFORMANCE EVALUATION

In the area of DBMS efficiency and performance evaluation, two aspects are considered: evaluation tools/techniques and evaluation results.

Reference 8-10 presents a survey of evaluation techniques. Although there are a number of analytical techniques under investigation, comparison of features and capabilities continues to be the most widely used technique for DBMS evaluation. As a result of the broad diversity of parameters influencing DBMS performance, simulation is expected to be the most useful tool for the near future.

Three of the most important performance parameters are storage overhead, CPU time, and data access time. Storage overhead is a ratio of storage consumed by the DBMS for retention and access of data (including the data, compressed or uncompressed) to storage consumed by the uncompressed data. Data access time reflects the non-CPU time consumed to retrieve requested data.

8.8.1 State of the Art in Efficiency/Performance

Storage overhead is primarily compressed disk overhead and varies widely from 50% to 400%. While data compression could be used to decrease storage overhead, it is presently used in only one or two DBMSs.

8.8.2 Trends in Efficiency/Performance

Since trends indicate utilization of large amounts of electronic storage in the future, storage overhead should not be limited to disks, as is typically done today. It is expected that any storage overhead that might be diminished because of hardware advances will be essentially reinstaced via incorporation of new or additional capabilities not previously available.

With the use of additional subsidiary processors, both CPU time and data access time consumed by DBMS will continue to decrease through 1985. It is anticipated that DBMS will exhibit the most significant decreases in CPU time in the 1980-1985 timeframe.

8.8.3 Projected Developments in Efficiency/Performance

Due to the recapturing of storage overhead by the addition of capabilities as described above, storage overhead is not expected to change appreciably by 1980, and perhaps may even increase by 1985. Due to projections in hardware technology, the period between 1980 and 1985 will exhibit the most significant decreases in CPU and data access times.

While other parameters may be of interest to specific installations, the above parameters are considered common to a wide variety of interests and applications for evaluation purposes. These and other identifiable DBMS performance parameters will be analyzed in more depth during subsequent program phases.

8.9 LARGE-CAPACITY MASS STORAGE SYSTEMS

8.9.1 Large-Capacity Magnetic Mass Storage Systems (MSS)

On-line, large-capacity magnetic mass storage refers to devices with capacities on the order of 10^{12} bits. A trillion bits is approximately equal to 2,900 reels of standard 1600 BPI tape, 3,500 IBM 2314 disk packs, or 1,000 IBM 3330 disk packs. The advantages of MSS over tape or disk storage devices are reduction of file storage space (on-line and off-line), automated management of file storage to eliminate manually induced errors, reduction of overall file storage cost after conversion of data to mass storage, improved security, better data availability, and simpler computer operations. Mass storage systems are used as extensions of main memory where access times up to 15 sec are adequate, for resident storage of control programs and/or compilers in or out of main storage in timesharing or multiprogramming environments, for storage of on-line data bases in a storage hierarchy scheme, and as a replacement for magnetic tape or disk storage systems.

There are basically three different methods for interfacing an MSS to the user CPUs: user direct, CPU staging, and MSS staging. The user direct interface allows application programs in execution to have direct access to the MSS either as an emulation device or via a special access method. The CPU staging interface uses a system-wide staging service utility to prestage files from MSS through the CPU to high-speed disks, thus decoupling application program execution from MSS access resulting in high channel activity and CPU overhead. The MSS staging interface removes the staging function from the CPU and places the burden on the MSS, thus providing more efficient staging with less contention but at increased cost and sophistication of hardware. The tradeoffs between the different interfaces involve file size, request rate, and cost. The user direct and CPU staging interfaces are cost/performance feasible with low request rates and small files. On the other hand, for high request rates or multiple, simultaneously processed files, MSS staging becomes more cost/performance effective over CPU staging as the average file size increases.

8.9.1.1 State of the Art in Large-Capacity Magnetic Mass Storage Systems - In industry today, there is a limited variety of commercially available mass storage systems. Since the technologies employed in these systems are considerably different, the state of the art is presented in terms of the specifications of the individual systems rather than in general terms as done throughout most of this report. Table 8.9.1.1-1 summarizes the characteristics of the four commercially available magnetic mass storage systems: SDC TBM II, IBM 3850, CDC 38500, and Calcomp 7110 Automated Tape Library (ATL).

The SDC TBM II mass storage system uses standard 2-in. transverse scan video tape. The system consists of parallel transport modules with transport drivers and data channels. The maximum system configuration is 32 transport modules, each with 9.2×10^{10}-bit capacity on two reels of tape. Data are recorded on dual channels in 1-in.-long addressable blocks on transverse tracks (189 TPI), providing 1×10^{6} bits per block. The error rate of 5×10^{-11} is achieved by using a dual-record scheme along with error correcting codes. This data-redundant format records data twice on tape, with a separation of 3/4 in. between records. Thus, to have an uncorrectable error, two tape defects would have to occur exactly 3/4 in. apart. The system allows up to six simultaneous accesses, which is equivalent to searching through 300 computer tapes per second. The TBM (terabit memory) system was originally developed by Ampex. SDC currently has exclusive marketing rights for the system.

The IBM 3850 is a hierarchical storage system that provides capacities of up to 472 billion bytes and offers disk-access capabilities, via an IBM 3330 disk system, by use of a virtual direct-access storage concept. The 3850 uses a flexible magnetic medium in individually spooled cartridges that are moved from a pair of 30-ft honeycomb-type library walls to and from magnetic recording stations using a helical rotary feed. Dual cartridge accessors are used to reach any cartridge in the library at search speeds of 99 in/sec. The cylindrical cartridge is 1.9 in. in diameter and 3.5 in. long and contains a tape

TABLE 8.9.1.1-1. STATE OF THE ART IN LARGE-CAPACITY MMS

SYSTEM CHARACTERISTIC	SDC TBM II	IBM 3850	CDC 38500	CALCOMP ATL
System Capacity (Gbyte)	Up to 362	35 to 472	16 to TBD	Up to 1,500
Subunit Capacity (Mbyte)	5,656/reel	50/cart.	8/cart.	150/reel
Average Access Time* (sec)	5	7 to 9	2.5 to 5	15 to 20
Data Rate (Kbytes/sec)	700/data chan.	806	806	1,200
Error Rate	5×10^{-11}	TBD	TBD	TBD
User Direct Interface	Yes	No	Yes	No
CPU Staging Interface	Yes	No	Yes	No
MSS Staging Interface	Yes	Yes	No	No
Virtual Volumes	No	Yes	No	Real volume
Virtual Files	Yes	No	Yes	No
Linear Density (bpi)	7,600	TBD	6,250	Up to 6,250
Areal Density (bpi^2)	1.5×10^6	3.36×10^5	3.33×10^5	5.6×10^4
Volumetric Density (bpi^3)	TBD	4.16×10^7	2.8×10^7	TBD
Maximum Data Set Size (Mbyte)	TBD	TBD	200	TBD
Tape Read/Write Speed (ips)	5	TBD	129	Up to 200
Recording Method	Transverse Rotary	Helical Rotary	Longi-tudinal	Longitudina
Recording Mode	NRZ/Asynchronous FM	GCR	GCR	NRZI/PE/GCR
Number of User CPUs	3	1 to 4	1 to 4	1 to 4
Number of Read/Write Stations	2 to 64	2	2 to 4	2 to 32
System Cost (m¢/bit)	0.1	0.07	0.2	0.08
Media Cost (m¢/bit)	6×10^{-4}	0.005	0.023	0.01
Operational Date	1972	1974	1974	1976 to 197

*Access Time To Submit

medium that is 2.7 in. wide and 770 in. long. The helical-scan transport design is used because of its ability to handle a wide medium without a large number of recording heads. Another advantage of the helical scan configuration is that the instantaneous data rate is independent of tape velocity, and a high-average data rate requires only low longitudinal tape velocity. The system also uses incremental step mode tape motion to improve error recovery time. The tape is moved forward one stripe width as the head is transversing the gap between stripes and is stationary on the stripe during reading or writing. Thus if an error is detected the tape remains stationary and is reread in an attempt at error recovery. Another feature of this system is that the tape is floated on a 10-μin-thick air cushion over the high-speed rotary head to eliminate head wear due to tape abrasion.

The CDC 38500 is similar to the IBM 3850 in that both use flexible media in a cartridge that is accessed from honeycomb-type library walls. The 38500, however, uses longitudinal recording with a positionable pair of nine-track heads that move across the width of the tape to one of eight positions, with a movement time of about 20 msec. The cartridge contains a 2.7-in.-wide, 150-in.-long ferric-oxide-coated magnetic tape (only 100-in. are used for recording). Data are recorded in 18-track data streams, with a total of eight addressable data streams per cartridge. The 38500 uses error detection and correction, on-line diagnostics, and the automatic ability to physically and logically reconfigure the MSS hardware in the event of a component failure to provide high reliability, availability, and maintainability.

The Calcomp Model 7110 ATL is the largest mass data storage and retrieval system commercially available as of 1977. It is approximately 100 ft long by 15 ft wide, with a capacity of up to 1.5 trillion bytes of data storage on either 7,775 thin line reels of magnetic tape or 6,122 standard half-inch tape reels. Approximately 15 sec are required for any ATL tape reel to be automatically retrieved and mounted on one of up to 32 tape drives in a system. The ATL can perform over 150 activities (mounts, dismounts, ejects, stores) per hour.

8.9.1.2 Trends In Large-Capacity Magnetic Mass Storage Systems - Mass storage systems (MSSs) are a rather young and still evolving storage technology. In the coming decade, MSS will have an ever-increasing customer base as users' desire for more storage grows, thus providing more momentum for further improvements in performance and price. Even though the initial entry price is high, the overall low cost per bit stored will ensure MSS to be a viable technology for on-line bulk storage through 1985. There is currently a growing trend to use MSSs in communication-based information systems because of the increasing cost of paper media and the desire by users to place massive amounts of data in a mode where it can be accessed interactively. Recent studies (Ref. 8-11) have shown that users do not mind an initial delay in order to have access to their data, making MSS very feasible. In various applications (e.g., sales forecasting, production and inventory control, customer information, and scientific and engineering calculations), the requirements for on-line data bases have grown more rapidly than the disk device technology. With the ever increasing appetite for on-line storage by industry and government, it is forecast that between 2,000 to 4,000 facilities worldwide will each require one or more trillion-bit systems by 1985. Figures 8.9.1.2-1 and 8.9.1.2-2 show magnetic mass storage capacity and cost trends, respectively.

Recent studies (Ref. 8-12) show that rarely used files and extremely large files (used less than bi-weekly) should remain on tape; files used many times a day should remain on disk; and all other files should be on mass storage devices. Between 50 and 70% of all tape files in a tape library and 85% of tape activity could be eliminated by using MSS. The following is a list of the type files that should be on mass storage devices to provide effective storage:

- Files smaller than 1 Mbyte used more often than once every 2 to 3 months and less than daily
- Files between 1 to 10 Mbytes used more often than 1 to 4 times a month and less than 1 to 3 times daily
- Files larger than 10 Mbytes used more often than 1 to 2 times weekly and less than 3 to 5 times daily.

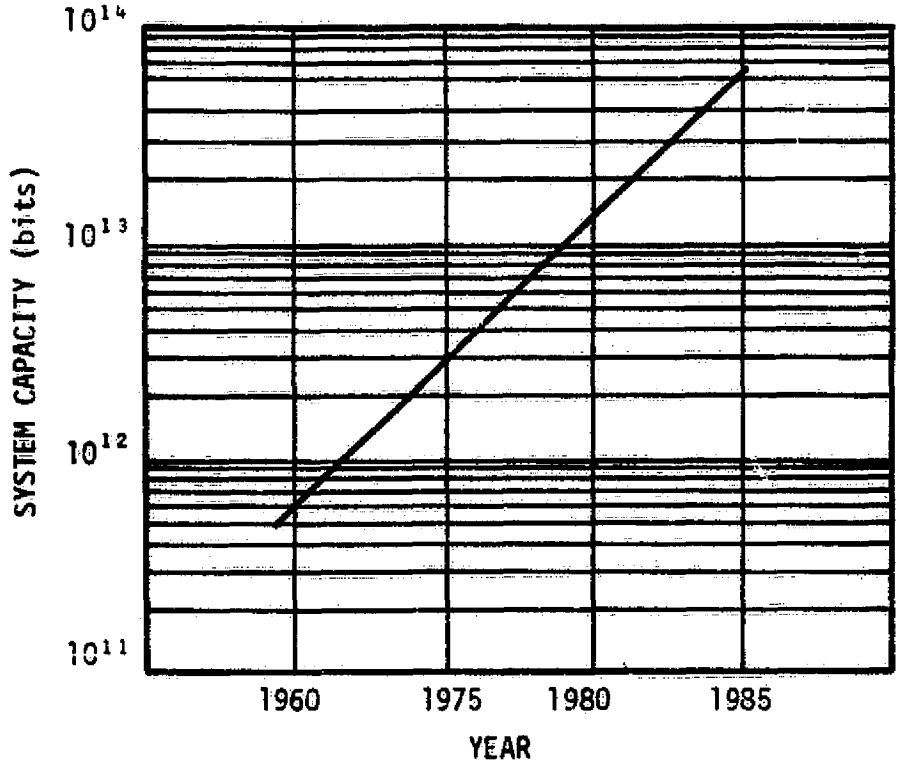


FIGURE 8.9.1.2-1. TRENDS IN MMS CAPACITY

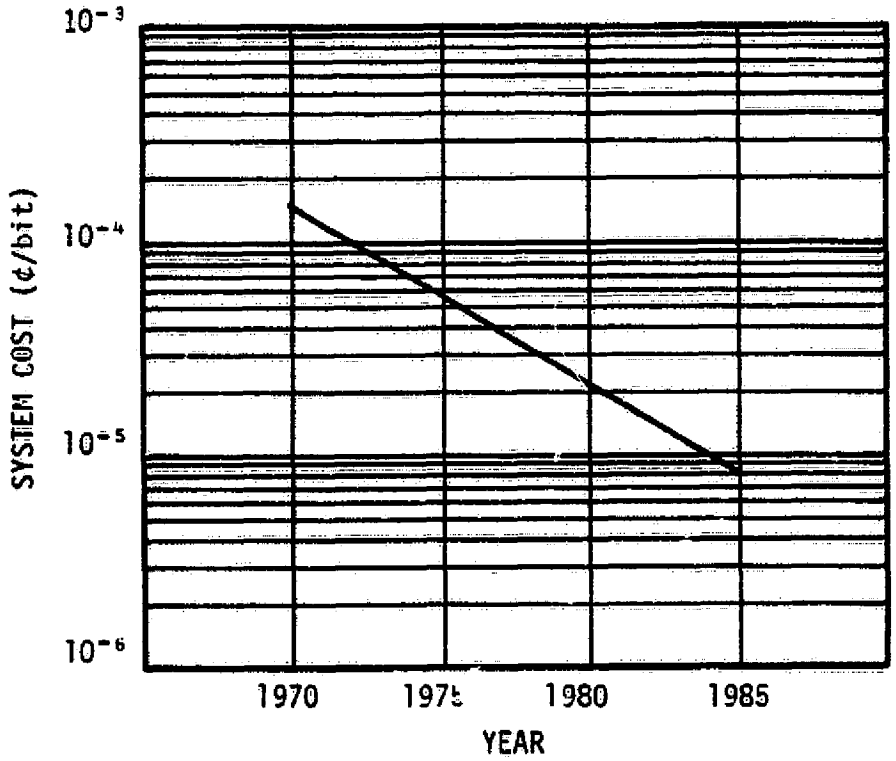


FIGURE 8.9.1.2-2. TRENDS IN MMS COST

8.9.1.3 Projected Developments in Large-Capacity Magnetic Mass Storage Systems - The trends in MSS over the last 5 years and the assessment of several authorities in the field are that magnetic mass storage systems will be the dominant large-capacity, on-line storage facilities in 1980 and 1985, with the characteristics given in Table 8.9.1.3-1. It appears unlikely that MSS will significantly reduce overall storage cost so long as it is mainly used to replace only part of the magnetic tape libraries and drives currently used by installations. Also, half-inch reel-to-reel magnetic tape is such an entrenched standard that it is difficult to foresee mass storage cartridges displacing tape within the next 10 years. As needs for larger-capacity MSSs develop, careful system planning will be necessary to produce not only larger and faster storage devices but also intelligent systems capable of serving multiple CPU configurations with modular capacity, throughput, and redundancy. Some current problems in MSS that will have to be solved are programming overhead that equals or exceeds the true access time of the device, the complicated addressing structure, excessive handling, time-dependent I/O, the costly conversion process of placing existing files on the MSS, and reliability.

By 1985, technology will support MSSs having up to 6×10^{13} bits on line; however, the predominant MSS will have a somewhat lower capacity. These systems will still use flexible magnetic media with a packing density of 10^7 bits/in^2. This is an order of magnitude higher than current systems and will be achieved without any loss in data reliability. There is currently no feasible technology approach to large-capacity MSSs that could do away with the use of physical motion or the use of continuous recording media. The basic storage hierarchy of future MSSs will include high-speed rigid disk (or electronic disk such as bubbles or CCD) with a greater amount of on-line storage under system control.

TABLE 8.9.1.3-1. PROJECTED DEVELOPMENTS IN LARGE-CAPACITY MAGNETIC MASS MSS, 1980-1985

SYSTEM CHARACTERISTICS	1980	1985
System Capacity	1.5×10^{13} bits	6×10^{13} bits
Subunit Capacity	8.0×10^{8} bits	1.6×10^{9} bits
Average Access Time*	2 to 4 sec	1 to 3 sec
Data Rate	1,500 Kbytes/sec	4,500 Kbytes/sec
System Cost	3×10^{-5} ¢/bit	8×10^{-6} ¢/bit
Areal Density	5×10^{6} bpi^2	1×10^{7} bpi^2

*Access time to submit

8.9.2 Optical Mass Storage

Although magnetic mass storage systems, such as discussed in the previous subsections will provide the capability for handling large-capacity on-line data base requirements of up to 5×10^{13} bits through the 1985 timeframe, there is increasing evidence that on-line data banks in excess of 10^{14} bits will be needed by that time. In fact, the charter and activities of NASA are such that, by the early 1980's, a need should exist for such a system for archival data storage. The present trends presented in this report (see Subsection 8.9.1.2) indicate that magnetic storage technology will not fulfill the need for very-large-capacity on-line mass storage ($>10^{14}$ bits) by 1985.

Although optical storage will not supersede magnetic storage in the foreseeable future, it is capable of competitive performance in storing large volumes of data and will most certainly be used to implement very-large-capacity mass storage systems by 1985.

The two basic optical storage methods are holographic and discrete. In the discrete systems, a single bit is stored as a hole, photographic dot, stored charge, etc., on some type of medium (film, metal-coated mylar, semiconductor plate, etc.). In the holographic systems, a group of bits are stored in either a one-dimensional or two-dimensional hologram. A series of holograms is written in rows and columns on the film (also called a hologram matrix).

Both discrete and holographic systems have been developed and fielded, but neither approach has yet gained a position that would favor it as the dominant technology of the future.

8.9.2.1 State of the Art on Optical Mass Storage - Some state-of-the-art characteristics for discrete and holographic optical storage systems are presented in Table 8.9.2.1-1. As seen in the table, state-of-the-art discrete storage systems currently offer higher storage densities than holographic systems, whereas the holographic systems provide higher data transfer rates. Some of the characteristics of each type of system are discussed in the following paragraphs.

Both the discrete and holographic storage systems use lasers for reading and writing information onto the film roll, strip, or fische. In the holographic system, data are recorded by first modulating the beam to create a spatial light image corresponding to the bit pattern to be recorded. The beam then passes through a Fourier transform hologram lens and is recorded on the film medium. The photosensitive medium is then developed and forms a hologram. Illumination of the film at the same relative orientation with the same wavelength laser light recreates the spectral image of the data, which can be sampled by a detector array to determine its contents.

The theoretical advantage to the holographic method was considered to be its fast (microsecond region) access time to on-line data. The standard definition of access time, though, is the time required to locate any bit of data that is stored by the system. If enough data are stored in the system, then there will be off-line data due to the physical limitations of any present optical storage medium. There appears to be no clear advantage in access time with this method unless the data are on-line, and then the access time depends on the packing density, the number of bits per unit area, of the medium. Physical limitations have shown that microsecond access times are not feasible, but one vendor has developed a system whose access time to 10^{12} bits of on-line data is 4 seconds. The data for this system are stored in a series of carrousels containing microfische.

The useful storage region for holographic media is limited by bookkeeping codes to 80% of the total area. Table 8.9.2.1-1 shows

TABLE 8.9.2.1-1. STATE OF THE ART IN OPTICAL MASS STORAGE

SYSTEM CHARACTERISTIC	DISCRETE STORAGE	HOLOGRAPHIC STORAGE
System Capacity	10^{12}	2×10^{11}
Media Capacity	2.5×10^{9}/105 in^2 strip	3×10^{7}/24 in^2 fische
Storage Density	25 Mbits/in^2	5 Mbits/in^2
Media	Roll or strip of film or metal coated mylar	Roll or strip of film
Average Access Time	10 sec	4 sec
Read Rate	4.0 Mbits/sec	40 Mbits/sec
Write Rate	NA	400 Mbits/sec
Media Cost	$\leq$$25 (strip)	NA
Error Rate	10^{-6}	$<10^{-5}$
Error Rate With Coding	10^{-13} to 10^{-14}	10^{-13} to 10^{-14}

that the best current packing density for holographic media is 10 Mbits/in^2. The Harris Corporation developed a holographic system with a packing density of 30 Mbits/in^2, but bookkeeping codes limited the useful storage region to 30% of the total area. Harris is not actively marketing this holographic system, since they have directed their efforts toward discrete archival data storage systems.

Another advantage of holographic storage is that it is less sensitive to media defects than discrete storage. Since the data are stored throughout the volume of the hologram, damage to a portion of the hologram does not destroy the data. It should be noted, however, that with the loss of a portion of the hologram comes a decrease in the signal-to-noise ratio and hence an increase in the bit error rate. A loss of 50% of the hologram area will correspondingly drop the signal-to-noise ratio by 50%.

In discrete storage, each bit of data occupies a distinct location on the surface of the medium. Successive data cells are interrogated by directing a focused beam of light on each data cell and examining the change in the resulting beam. Optical beams can be focused down to an area less than a square micron, implying a potential storage density of one bit per square micron, or 600 Mbits/in^2. Current discrete storage systems provide storage densities of up to 50 Mbits/in^2. Typical access time for the discrete systems is approximately 10 sec for on-line data.

Unlike the holographic system, a loss of a portion of the digital record means an irretrievable loss of that segment of data. Further, dust particles or other finite-sized contaminants can obscure individual data sites due to their small size (3 to 5 μm); the smaller the data site, the easier the obstruction may be. This means that the greater the packing density, the greater the potential bit error rates. NASA Ames is currently using a 64,16 Fire code in conjunction with their Precision Instruments Unicon system for error control. Redundant recording is also used with the system to reduce the 10^{-9} single copy error rate to approximately 10^{-18} errors per bit.

8.9.2.2 Trends and Projected Development in Optical Mass Storage - The major general trends in optical storage are toward increased storage density and reduced cost per bit of storage. Figure 8.9.2.2-1 illustrates how the storage density of optical storage devices is expected to evolve through 1990. By 1985, it is expected that holographic storage may achieve operational storage densities that are equal to or greater than the storage densities in discrete storage systems. The expected storage density by 1985 is 120 Mbits/in^2 for each layer in the medium. By recording five layers deep, 600 Mbits/in^2 of recording media could be achieved by 1985.

Because of the mechanical nature of both the optical and the holographic systems, data access times are not expected to improve significantly by 1985. Table 8.9.2.2-1 presents the characteristics projected for 1985 for both the discrete and holographic systems.

It is expected that by 1980 archival mass storage systems will be available that are based on one or more of the low-cost technologies currently being introduced as home video disk players. Several companies, including Hitach; Philips, MCA, Matsushita, Magnavox, RCA, and Sony have either already announced products or their intention to do so. Although these systems are all analog, laboratory digital systems based on similar technologies have demonstrated storage densities of 300 Mbits/in^2.

Recent advances have brought about erasable holographic media. If this technology advances as vendors claim, then holographic archival storage systems will have a read-write capability, whereas discrete systems will be read-only.

In addition to optical storage systems that employ visible lasers as the mechanism for reading and writing, some companies are beginning to research alternative storage technologies that employ either electron beams, ion beams, or X-rays for reading and writing into data storage cells with submicron dimensions. Submicron storage densities will probably be needed in order to develop a practical mass storage system with a total on-line storage capacity in excess of 10^{15} bits.

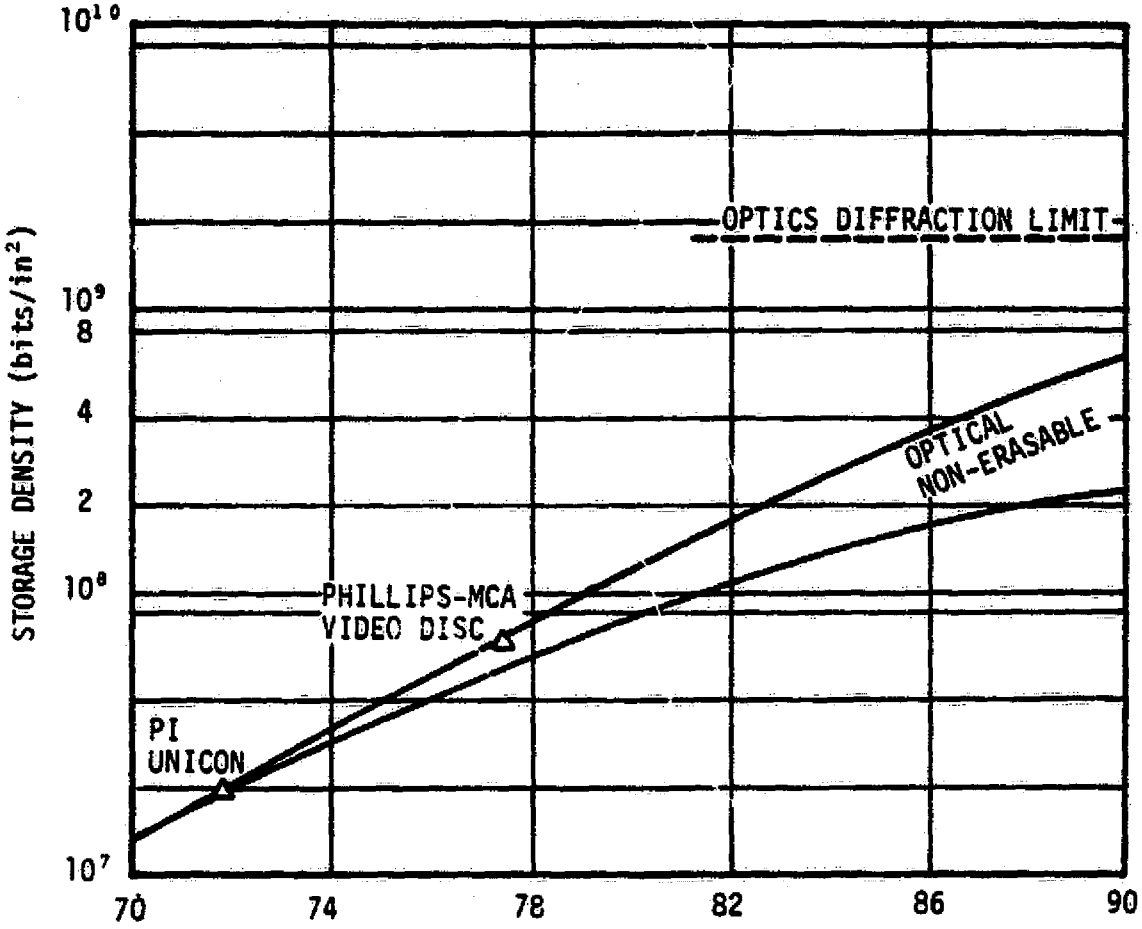


FIGURE 8.9.2.2-1. OPTICAL STORAGE DENSITY TRENDS

TABLE 8.9.2.2-1. 1985 PROJECTION FOR OPTICAL STORAGE TECHNOLOGY

SYSTEM CHARACTERISTIC	DISCRETE STORAGE	HOLOGRAPHIC STORAGE
System Capacity	10^{14}	1.2×10^{14}
Storage Density	250 Mbits/in^2	120 Mbits/in^2*
Average Access Time	10 sec	4 sec
Read Rate	10 Mbits/sec	100 Mbits/sec
Write Rate	10 Mbits/sec	600 Mbits/sec

*Possibly increased to 600 Mbits/in^2 by recording 5 layers deep.

One such research program, currently being funded by ARPA, is the development of an Advanced Archival Memory System for applications requiring at least 10^{14} bits. The system, which is being developed by General Electric, uses a derivative of GE's E-beam technology (see Subsection 7.2.2.1) to write permanent data onto semiconducting substrates with storage densities of up to 6×10^{10} bits/in^2. Each substrate, which would measure approximately 10 by 10 cm, would contain approximately 10^{12} bits. As the system is currently envisioned, it would consist of up to 10^3 on-line substrates plus a warehouse (or library) for off-line data storage. The system would provide multiple read and write stations with each read station capable of reading at a 10-Mbit/sec rate. The Avionics Laboratory at Wright Patterson AFB is monitoring the program for ARPA and should be contacted for further information.

REFERENCES - SECTION 8

8-1 "Study of Systems and Techniques for Data Base Management", Teledyne Brown Engineering, March 1977

8-2 "Introduction to Input/Output Requirements Language (IORL)", Teledyne Brown Engineering, March 1977

8-3 J. Martin, Principles of Data-Base Management, Printice-Hall, Inc., Englewood Cliffs, N.J., 1976

8-4 NBS, "CODASYL Data Description Language", U.S. Department of Commerce, January 1974

8-5 Kenneth J. Thurber and Peter C. Patton, Data Structures and Computer Architecture, D. C. Heath and Company, 1977

8-6 M. E. Senko, "Data Structures and Data Accessing in Data Base Systems Past, Present, Future", IBM Systems Journal, Vol. 16, No. 3, 1977

8-7 "Data Base Directions -- The Next Steps", Data Base, ACM Special Interest Group on Business Data Processing, Vol. 8, No. 2, Fall 1976

8-8 Baum and Hsiao, "Data Base Computers -- A Step Toward Data Utilities", IEEE Transactions on Computers, Vol. C-25, No. 12, December 1976

8-9 A Survey of Existing Techniques for Evaluating Data Base Management Systems, CSC Field Services Division, Huntsville, Alabama, May 1975

8-10 "A Buyer's Guide to Data Base Management Systems", Datapro Research Corporation, May 1976

8-11 G. C. Bacon, "Mass Storage Workshop Report", Computer, pp. 64-65, October 1974

8-12 L. Shirley, "The Spectrum of MSS Effectiveness", IEEE Spring 1976 Compcon, pp. 50-52.

BIBLIOGRAPHY - SECTION 8

"Evaluation of Data-Base Management Systems", ACM Computing Surveys, Vol. 8 No. 1, March 1976

R. Turn and W. H. Ware, "Privacy and Security Issues in Information Systems", IEEE Transactions on Computers, Vol. C-25, No. 12, December 1976

R. M. Curtice, "The Outlook for Data Base Management", Datamation, Vol. 22, No. 4, April 1976

V. Chvalovsky, "Anything New in Data Base Technology?", Datamation, Vol. 22, No. 4, April 1976

C. T. Johnson, "The IBM 3850: A Mass Storage System with Disk Characteristics", Proceedings of the IEEE, Vol. 63, No. 8, pp. 1166-1170, August 1975

J. Beeler, "Automated Library Holds 8,351 Reels, Cuts Labor Costs", Computer World, pp. 49, 57, October 17, 1977

H. R. Howie, Jr., "More Practical Applications of Trillion-Bit Mass Storage Systems", IEEE Spring 1976 Compcon, pp. 53-56, 1976

M. Wildman, "Terabit Memory Systems: A Design History", Proceedings of the IEEE, Vol. 63, No. 8, pp. 1160-1165, August 1975

A. S. Hoagland, "Magnetic Recording Storage", IEEE Transactions on Computers Vol. C-25, No. 12, pp. 1283-1288, December 1977

J. P. Harris, et al., "The IBM 3850 Mass Storage System: Design Aspects", Proceedings of the IEEE, Vol. 63, No. 8, pp. 1171-1176, August 1975

R. N. Nelson and T. H. Otten, "High Capacity Optical Data Storage and Retrieval Systems", Proceedings of International Telemetering Conference, pp. 348-355, September 1976

A. M. Bandos, et al., "Multigigabit Per Second Digital Data Recording", Electro-Optical Systems Design Conference, November 1975

9. INFORMATION DISTRIBUTION ELEMENTS

The information distribution elements covered in this section are the electronic distribution elements presented in Figure 9-1. The electronic distribution elements may be broken down into telecommunication hardware and telecommunication networks. The networks are further divided into satellite networks and terrestrial networks, with the latter being categorized as analog, digital, or value-added. The telecommunication hardware can be broken down into a number of devices as shown. Communications terminals are covered as information presentation elements in Section 10.

The state of the art in telecommunication network technology is presented in terms of the capabilities offered by the specific types of networks. It should be noted that in many cases these networks are by no means pushing the state of the art in terms of what is achievable, and the availability of network services is not directly tied to projected technology. They are instead what is practical from an availability and cost viewpoint. There are a number of items, however, that offer the potential for increasing the network capabilities, and these items are addressed in terms of their applications to offering a lower-cost wideband data transfer capability.

Telecommunication hardware is addressed in terms of present capabilities. Again, many units exist that do not pressure the state of the art, but they do provide a cost-effective system when implemented with existing hardware. The telecommunication hardware is also analyzed from the standpoint of a distributed network, with emphasis given to the impact on the hardware and/or network.

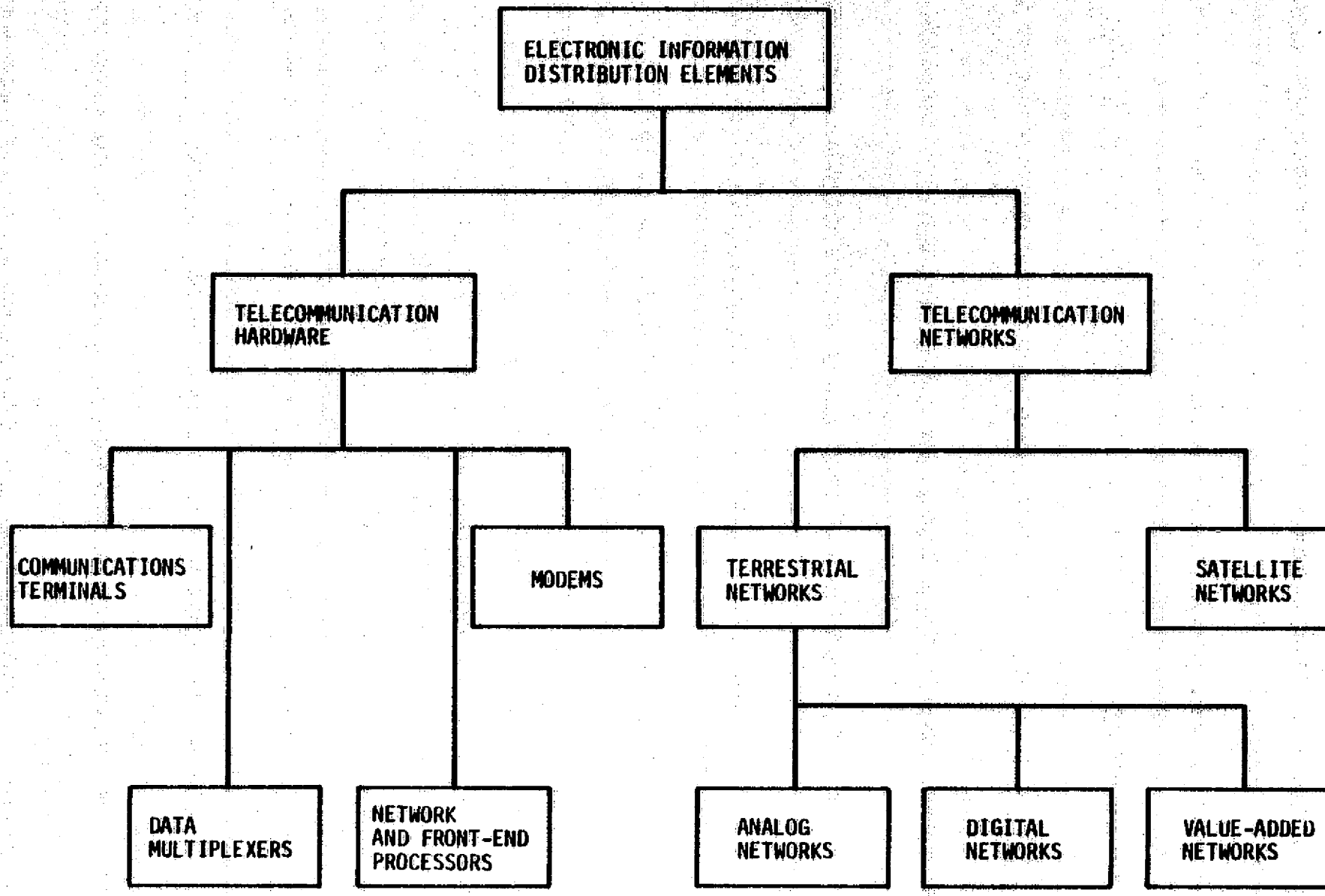


FIGURE 9-1. ELECTRONIC INFORMATION DISTRIBUTION ELEMENTS

9.1 TELECOMMUNICATION NETWORKS

Telecommunication networks may be characterized according to the media used for their implementation and the characteristics they possess. For purposes of technology assessment, it seems appropriate to characterize the networks as either terrestrial or satellite. The terrestrial networks may be further defined as analog, digital, or value-added.

The capabilities and the technology are presented in terms of network performance and cost, irrespective of whether the network is common carrier, specialized common carrier, or private. A distinction is made for instances where a carrier offers a special performance advantage.

Terrestrial telecommunication networks will continue to predominate in the transmission of data in the early to mid-1980's. Long-haul links will employ satellite channels, but most users will tend to interface with these channels through the regular telephone system. The use of small satellite terminals will see increasing popularity during the 1980-1985 timeframe, particularly for the reception of wideband data. The use of these small terminals for interactive data communications will also increase, but this use will be somewhat restricted as a result of interface problems.

9.1.1 Terrestrial Networks

Terrestrial networks are typically classified as either analog or digital. Both analog and digital network services are available from AT&T and a number of independent telephone companies (e.g., GTE, UTS) and from specialized common carriers (e.g., MCI Telecommunications Corporation, Nebraska Consolidated Communications Corporation, Southern Pacific Communications Company, and a number of other companies). Value-added networks, while generally digital in nature, are covered as a third area due to their uniqueness of service and implementation. The characteristics of each of these networks are presented in terms of services and performance provided by the network.

The general complexity and the constantly changing status of carrier tariffs excludes an in-depth analysis of the subject, but several cases are analyzed for data links between Huntsville, Alabama, and White Sands, New Mexico, the site of the TDRSS ground station. Table 9.1.1-1 presents the data for these links. Slow-speed land lines are available to the customer on a dial-up basis, with the maximum data rate being 4.8 Kbits/sec. Leased voice-grade lines are available that have a capability of handling 9.6 Kbits/sec. Since the carrier used digital transmission techniques, a 56-Kbit/sec digital line is available for approximately the same price. The mileage charge is the driving factor for the cost of either line, and the charge is approximately $12,000/month in both cases. Terrestrial lines from White Sands are currently limited to the 56-Kbit/sec links, but an increased data rate can be accommodated by using parallel lines with the price per line dropping slightly as the number of lines increases. Equipment to accommodate 1.544-Mbit/sec digital service will be added to the White Sands facility at a future date, but pricing information on this link is currently unavailable. The cost of similar 1.544-Mbit/sec links, with distances comparable to that from Huntsville to White Sands, currently runs in the range of $60K to $70K per month.

TABLE 9.1.1-1. COST AND AVAILABILITY OF TERRESTRIAL DATA LINES FROM HUNTSVILLE TO WHITE SANDS

TYPE CIRCUIT	BIT RATE	COST
Direct Distance Dialed	4.8 Kbps	$0.35 per minute
Leased Voice Grade	9.6 Kbps	$12K/month
Leased Medium Rate Digital	56 kbps	$12K/month
Leased High Rate Digital	1.544 Mbps	$60K to $70K/month

Analog networks are the primary networks in use at the present time, but trends and technological developments are mainly in the area of digital networks. Development activity is also being applied to satellite relay networks and value-added networks. The major developments for analog networks are in the switching hardware and the techniques for interfacing to the network. The analog networks will play a decreasing (although still major) role in the telecommunications networks of the 1980's, with their replacement being slow due to the amount of equipment currently in service. The other networks (e.g., digital, value-added, and satellite) will grow by comparison, offering services unavailable on the analog network. The major emphasis of this section is therefore placed on the other networks, especially since they lend themselves readily to data transmission.

9.1.1.1 State of the Art in Analog Networks - The principal telecommunications networks available within the United States today are of the analog type. Standard lines, whether from the public switched (dial-up) network or a hardwired leased network, offer a 3-kHz bandwidth extending from 300 to 3,300 Hz. Table 9.1.1.1-1 presents a representative list of the types of service available via analog networks. The typical data rates available for each type of line are also presented. The companies providing these lines may use either cable or terrestrial or satellite microwave links (or a combination) since network transparency is the major requirement of the end user.

The primary use of the analog network continues to be for voice, but the transmission of digital data over analog lines has grown rapidly within the past several years and is expected to grow at a faster rate in the future. The transmission of digital information over analog networks requires the use of a modem (as described in Section 9.2.1). Custom-tailored analog services are available (e.g., wideband equalized lines and control lines). The carrier assesses user requirements and provides the facilities and services to satisfy that specific requirement. Of course the cost for this type of service goes beyond the tariffs for standard services.

TABLE 9.1.1.1-1. TYPICAL ANALOG NETWORK SERVICES WITHIN THE UNITED STATES

TYPE SERVICE	BANDWIDTH (kHz)	MAXIMUM BIT RATES (bits/sec)	SUPPLIER	COMMENTS
Dial-Up Lines				
Subvoice Grade		To 180	Western Union	
Direct Distance Dialed (DDD) Network	3	To 4,800*	AT&T, GTE, and others	
Wide Area Telephone Service (WATS)	3	To 4,800*	AT&T	Leased
BEX Service		To 38,400*	Western Union	Switched service. Available in certain areas only.
Leased Lines				
Subvoice Grade		To 180	AT&T and WU	
Voice Grade Lines	3	To 9,600	AT&T, WU, and Specialized Common Carriers	C1, C2, and C4 conditioning available
Wideband	To 240	To 230,400	AT&T and Specialized Common Carriers	Available in Groups (12 voice-grade lines), Super Groups (60 voice-grade lines), and Master Groups (240 voice-grade lines)

*Dependent on modems. Higher rates may be achieved.

9.1.1.2 State of the Art in Digital Networks - Digital networks offer a straightforward approach for transmission of digital data, but they present some unique problems not encountered in analog networks. The digital networks use signal regenerators to provide a high-quality digital signal with low error rates. These regenerators, spaced periodically along the transmission line, sample the incoming bit stream and produce a corresponding digital bit stream free of noise, crosstalk, and distortion. Since the hardware used in these digital transmission systems is compatible only with digital signals, any analog data that are transmitted must be digitized to correspond to the network's digital format. A CODEC (coder/decoder) is a device commonly used to convert voice data to a form suitable for digital transmission. A typical CODEC samples the voice data at a rate of 8,000 times per second, with each sample coded into 8 bits. Thus the bit rate for one voice channel is 64 kbits/sec. The CODEC functions are typically implemented on one or two IC chips, with output rates programmable between 64 kbits/sec and 3 Mbits/sec. The cost of these chips runs from $5 to $10, depending on the functions available. The 64-kbit/sec CODEC output is somewhat high when compared with the 3-kHz analog voice signal. Lower bit rates can be used, but the reconstructed analog signal degrades in quality for rates much below 64 kbits/sec. Several approaches have been considered for lowering the bit rate to a more manageable level. Some of these signal processing techniques rely on the redundancy of normal speech, while others use adaptive processing to take advantage of the pauses found in most conversations. Even without the use of signal processing to lower the data rates, the telephone company saves significant amounts of money by using digital switching equipment. As compared with existing analog equipment, digital switching equipment requires fewer buildings to house equipment, has lower power consumption, makes more efficient use of transmission trunks, and allows more economic expansion. To make full use of these benefits, telephone companies are increasing their digital networks at the rate of tens of millions of circuit miles per year. These digital transmission networks allow optimum transmission of both digital data and analog data that have been digitally encoded.

Digital data networks are growing at a rapid pace, and a number of specialized common carrier and private networks offer digital services. The digital network that best represents the state of the art is AT&T's Dataphone Digital Service (DDS). This network is currently expanding to offer service to more than 100 cities, as listed in Table 9.1.1.2-1. The services offered by DDS range from lines of 2.4 kbits/sec up to 1.544 Mbits/sec. Additionally, AT&T is implementing a Dataphone Switched Digital Service (DSDS) to allow switched digital lines operating at speeds up to 56 kbits/sec. It is anticipated that these lines will be used for batch transmission services with the average connection time being approximately 3 sec. Additionally, Western Union has begun a digital service at 9.6 and 56 kbits/sec, with plans to offer 2.4- and 4.8-kbits/sec service in the future. The Western Union system will include service to approximately 50 cities.

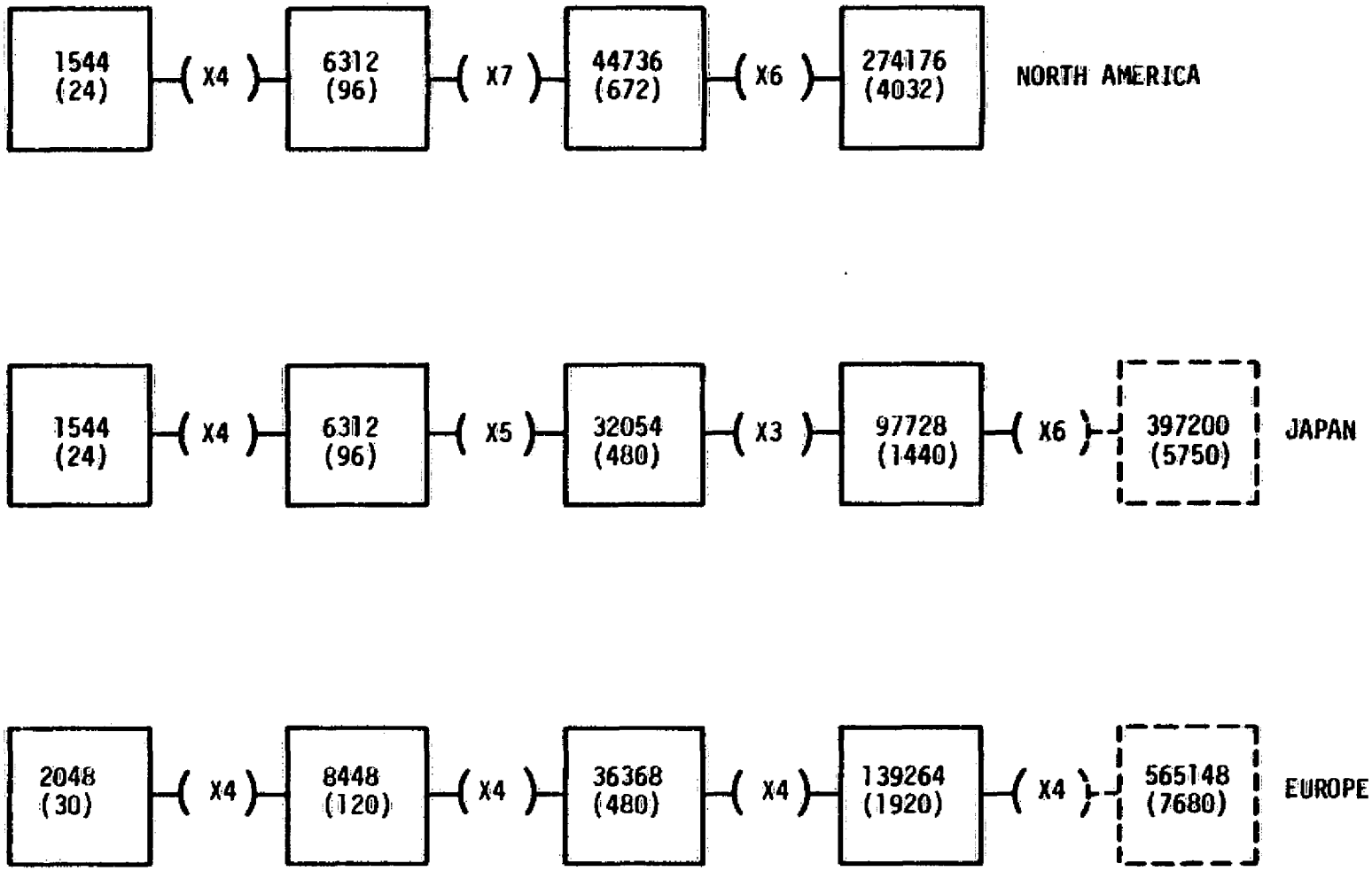
In implementing their DDS, AT&T multiplexes each user's data onto a T1 (1.544-Mbits/sec) line and this in turn is eventually multiplexed onto the higher-data-rate links via the digital hierarchy of the Bell system. The digital hierarchies for North America (the Bell system), Western Europe, and Japan are presented in Figure 9.1.1.2-1. Each of the three hierarchies is currently implemented through the lower four levels. The fifth level (Japan and Europe) has not been implemented at this time.

The full realization of an efficient digital network is dependent on the equipment used in the network. Development efforts are currently underway to increase the throughput of the digital multiplexers used to interleave the data coming from lower speed lines. Efforts are also being applied to increase the performance of the existing terrestrial microwave links. Improved cables and waveguides are being developed for accommodating the higher data rates, and optical fibers are being tested as a means of linking high traffic-switching stations.

Multiplexer technology is extremely important in the realization of a high-speed digital network. While ground-based telecommunications

TABLE 9.1.1.2-1. CITIES TO BE SERVED BY DDS

Boston, New York, Washington, D.C., Philadelphia, Chicago
Baltimore, Pittsburgh
St. Louis
Cleveland
Newark
Detroit
Kansas City
Dallas
New Haven, Denver
Hartford, Los Angeles
Atlanta
Milwaukee, San Francisco
Houston, Miami
Portland, Minneapolis
Camden, Seattle, Indianapolis, Mercerville (Trenton), Salt Lake City, White Plains
Wilmington, Memphis, Springfield, Massachusetts
Inglewood, California, East Bay, California, Omaha
Anaheim, California, Mountain View, California, Albany, Orlando, Buffalo
Charlotte, Phoenix, Columbus, Akron, Oklahoma City, Syracuse, Rochester
Des Moines, Sacramento, Dayton, San Diego
Tulsa, Cincinnati, Providence, Harrisburg, New Orleans
Birmingham, Toledo, Worcester, Allentown
Jacksonville, Louisville, Huntsville, Raleigh, Nashville
Greenville, Tampa, Richmond, Youngstown
Spokane, Reno, Shreveport, Colorado Springs
Greensboro, North Carolina, Norfolk, Fresno, Knoxville South Bend
Wichita, Ventura, Albuquerque, Las Vegas, Lansing
Fort Wayne, El Paso, Lincoln, Chattanooga, Reading
Grand Rapids, Peoria, San Antonio, Baton Rouge, Madison
Flint, E. Moline, Little Rock, Roanoke, Cedar Rapids



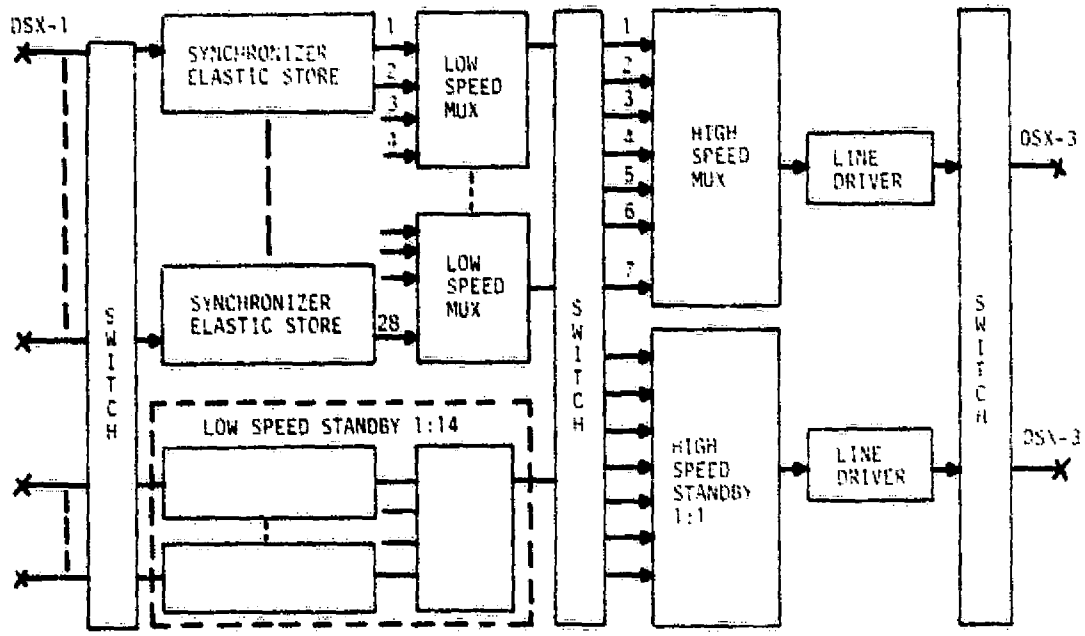
NOTES:

- FIRST NUMBER IS BIT RATE IN kbits/sec
- NUMBER IN PARENTHESIS IS NUMBER OF 64 kbits/sec TIME SLOTS
- PROPOSED - - - - -

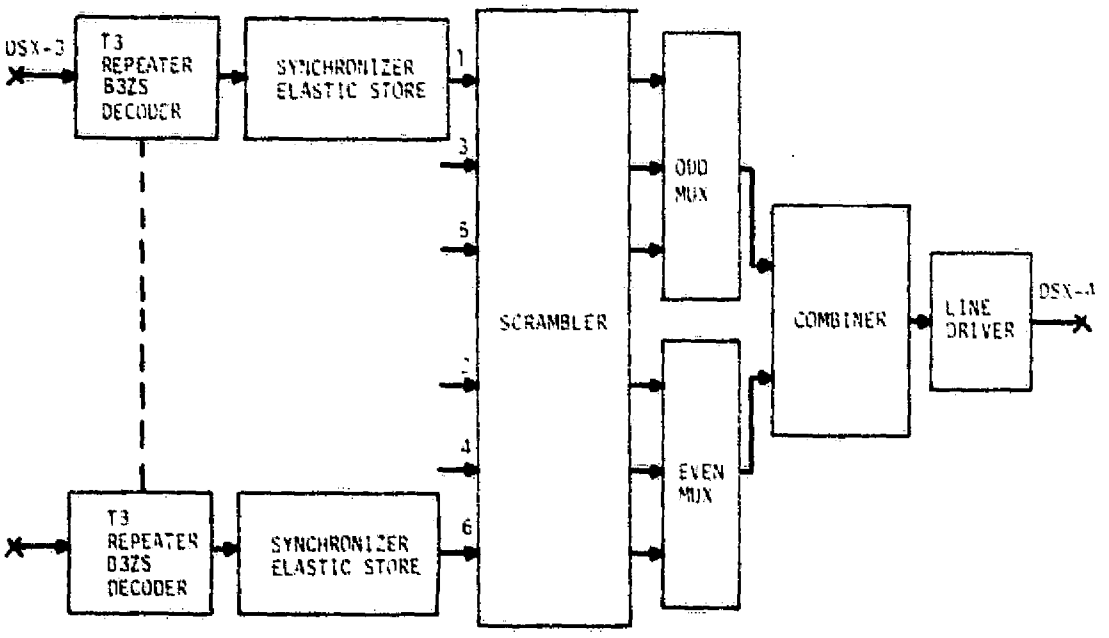
FIGURE 9.1.1.2-1. DIGITAL COMMUNICATION NETWORK HIERARCHY

multiplexing equipment is not severely limited in terms of power or size (although the lower the better), system reliability and cost are major factors of consideration. The state of the art in operational telecommunications multiplexers is the 274.176-Mbit/sec system currently being used in the Bell system. This M34 (level 3 to 4) digital multiplexer was placed in operation between Newark, New Jersey, and New York City in early 1975 on a test basis. The performance of the system has been excellent and it is now a standard operating system. Bell has implemented these high-rate multiplexers in other areas and currently has six in operation with plans to install more as the traffic needs justify them. In operation, 28 of the 1.544-Mbit/sec DS1 lines are combined into a single 44.736-Mbit/sec (DS3) line using an M13 (level 1 to 3) multiplexer. The outputs of six of the M13 units are then combined using the M34 multiplexer to obtain the 274.176-Mbit/sec signal. Figure 9.1.1.2-2 presents the block diagram configurations for both the M13 and M34 multiplexers. Multiplexing in the M13 unit is performed on two levels. The first level combines four of the 1.544-Mbit/sec lines into a 6.312-Mbit/sec line. Seven of these lines are combined to obtain the 44.736-Mbit/sec output. The low-speed multiplexers have a redundancy of 1 backup for every 14 on-line units, while the high-speed multiplexer has a redundant unit for each operational unit. The six 44.7-Mbit/sec lines are fed into the M34 multiplexer and each line is scrambled using a pseudo-random bit stream to control the transition density of the output signal. The odd lines and the even lines are processed individually using ECL at a speed of 2 nsec. These two 137-Mbit/sec lines are then combined using 1-nsec logic. By combining these two streams near the multiplexer output, the 1-nsec logic is limited to approximately 4% of the total number of IC packages. The characteristics of the multiplexer are presented in Table 9.1.1.2-2 and the framing and stuffing performance is given in Table 9.1.1.2-3.

The transmission network for the fourth level of hierarchy in the Bell system is designated as the T4 digital transmission system. The system is capable of transferring the data from 168 T1 (1.544-Mbit/sec)



a. M13 Multiplexer Configuration



b. M34 Multiplexer Configuration

FIGURE 9.1.1.2-2. M13 AND M34 MULTIPLEXER CONFIGURATIONS (Ref. 9-1)

TABLE 9.1.1.2-2. M13 AND M34 MULTIPLEXER CHARACTERISTICS (Ref. 9-1)

MULTIPLEXER CHARACTERISTIC	M13		M34
	LS-MUX	HS-MUX	
Output Rate (Mbits/sec)	6.312	44.736	274.176
Stability (ppm)	±20	±20	±10
Frame Length	294	680	196
Superframe Length	1,176	4,760	4,704
Average Maximum Reframe Time (μsec)	7,050	1,280	120
Nominal Stuff Rate (Hz)	1,800	3,655	27,429
Maximum Stuff Rate (Hz)	5,367	9,398	58,286
Stuff Ratio (%)	33.5	38.9	47.1

TABLE 9.1.1.2-3. M13 AND M34 MULTIPLEXER FRAMING AND STUFFING PERFORMANCE (Ref. 9-1)

PERFORMANCE / ERROR RATE	MEAN TIME BETWEEN MISFRAMES (sec)		MEAN TIME BETWEEN IMPROPER DESTUFFS PER CHANNEL (sec)	
	M13	M34	M13	M34
10^{-6}	2×10^{12}	1.2×10^{9}	2.2×10^{7}	4.7×10^{9}
10^{-3}	3.4×10^{3}	4.5×10^{1}	2.4×10^{1}	1.7×10^{3}

lines along a pair of coaxial tubes designated as the T4M line. Regenerators are required on the T4M line at distances not to exceed 5,700 ft. A typical maintenance span for the line is less than 111 mi. in length, with the total system length limited to 500 mi. Plans for implementation of the DS4 carriers call for use of WT4 millimeter-waveguide systems and the DR-18 digital radio system.

The development of high-rate digital transmission networks is closely linked with the development of optical fiber transmission links. The fiber-optic transmission systems offer several improvements over coaxial cable and metallic wire. Main advantages include greatly increased bandwidth, lower loss, small diameter cables, decreased weight, elimination of crosstalk, and immunity of RF and inductive interference. Additionally, the price is competitive with wire/coaxial systems, and the cost is predicted to decrease as the manufacturing techniques are refined.

Fiber-optic transmission systems consist of three basic segments: the transmitter, the optical fiber, and the receiver. Most transmitters (also called light launchers) used today are either of the Light Emitting Diode (LED) type or the semiconductor laser type. LEDs offer advantages of low cost, long life, and simple modulation, but have the disadvantages of low optical output and lower efficiency. Semiconductor lasers, however, offer high power output and spectral purity but are plagued by longevity problems and the need for temperature control.

The optical fibers fall into three general categories: step-index multimode, graded-index multimode, and single-mode. The step-index multimode fiber is the simplest to fabricate but exhibits limited bandwidth due to intermodal dispersion of the propagating pulse. Graded-index fiber reduces this dispersion by radially varying the refractive index within the core. Intermodal dispersion is totally eliminated in the single-mode fiber, giving it extremely wide bandwidth capability. The small diameter of the core (5 to 10 μm), however, limits the light source to a laser and places extremely tight tolerances on the connections

between fibers. Receivers generally are either a silicon PIN diode or an avalanche photodiode. The PIN diodes exhibit a short response time, high quantum efficiency, low capacitance, and very low noise equivalent power. Avalanche photodiodes exhibit a high gain-bandwidth product, very fast risetimes, and low capacitance. They also function to increase signal-to-noise ratio when the system is limited by preamplifier noise. Unfortunately, these diodes are expensive and are subject to drift with temperature. Data on the characteristics of both transmitting and receiving devices are presented in Table 9.1.1.2-4. Additionally, data on fiber-optics is presented in Section 4.1.1.2.2 under fiber-optic data buses for onboard data handling.

The first major test of optical fibers in a telecommunications environment occurred in Atlanta in 1976 when Bell Laboratories installed two fiber-optic cables into existing underground ducts. These tests were intended to simulate the environment of a typical urban telecommunications application. The system used GaAs injection lasers as light sources and avalanche photodiodes as receivers. Several other experimental systems followed, each a simulation of operations. In May 1977, Bell Laboratories began a test in Chicago using the fiber-optic lines to carry actual telephone traffic. The Chicago installation uses a 24-fiber cable to link two central offices 1.5 mi. apart. Additionally, the cable links into an office building located between the two phone company locations. In 60,000 hr of operation, the system had three electronic circuit malfunctions and one failure of the gallium-aluminum-arsenide laser. The cable operates with a data rate of 44.7 Mbits/sec, the equivalent of a T3 digital transmission line. Results of that test are promising, with the cable loss averaging 8.5 dB/km for the cable and splices combined.

In Great Britain a similar fiber-optic system was put into service in mid-1977 by Standard Telephones and Cables (a subsidiary of ITT). The British system uses about 9 km of cable with two repeaters in the system at 3-km intervals. The network operates at 140 Mbits/sec,

TABLE 9.1.1.2-4. CHARACTERISTICS FOR ACTIVE DEVICES USED FOR FIBER-OPTIC TRANSMISSION SYSTEMS

DEVICE	USE	ADVANTAGES	LIMITATIONS
Light Emitting Diode (LED)	Transmitter	Low cost Long life Simple modulation	Low optical output Low efficiency
Laser Diode	Transmitter	High power output Spectral purity	Short life Need for temperature control
PIN Diode	Receiver	Short response time High quantum efficiency Low noise equivalent power	Preamplifier noise
Avalanche Diode	Receiver	High gain-bandwidth product Very fast risetime Low capacitance	Cost Need for temperature control

the fourth level of hierarchy within the European digital network. The system uses gallium-aluminum-arsenide lasers for light sources. Other optical fiber systems are being tested, but most are low-capacity with data rates of only a few megabits per second.

One physical limitation encountered in optical fiber communications is that the allowable pulse rate is inversely proportional to the fiber length. As the data rates increase, regeneration devices must be spaced closer within the transmission link. Table 9.1.1.2-5 presents some cost/performance characteristics for state-of-the-art optical fibers. Current step-index multimode cable (Ref. 9-2) is capable of operation at 10 Mbits/sec/km. Graded-index multimode fiber has a transmission rate of 150 Mbits/sec/km for LED drivers and 500-Mbits/sec/km for laser drivers. Single-mode fibers have a transmission rate in excess of 5 Gbits/sec/km, but critical splicing tolerances have limited these cables to laboratory environments.

TABLE 9.1.1.2-5. COST/PERFORMANCE CHARACTERISTICS FOR STATE-OF-THE-ART OPTICAL FIBERS

FIBER TYPE	ACHIEVABLE RATE	RELATIVE COST	COMMENTS
Multimode Step-Index	10 Mbits/sec/km	Low	Narrow bandwidth
Multimode Graded-Index	150 Mbits/sec/km	High	LED driver
	500 Mbits/sec/km		Laser driver
Single-Mode Step-Index	5 Gbits/sec/km	Moderate	Difficult to splice

9.1.1.3 State of the Art in Value-Added Networks - Value-added networks, also known as packet-switching networks, offer a very unique method for accomplishing digital data transfer. The first large-scale experimental packet-switching network was ARPANET, a project originated by the Advanced Research Projects Agency. Today such networks exist in a number of countries of the world and connections between some networks are available. Common carrier networks include TELENET and TYMNET in the United States, DATAPAC in Canada, TRANSPAC and CYCLADES in France, EPSS in the United Kingdom, the European Integrated Network across Europe, and TIDAS in Sweden. Additionally, a number of other networks are being considered.

In the concept of packet switching, data are transferred on a block, or "packet", basis rather than on a bit or message basis. The network receives the data from the originator and divides the data into packets. Each packet is routed along the diversified network from node to node, with each packet taking the route having best availability. The packets are passed to the next node, with several packets moving across parallel paths of the network. As each packet reaches the destination node, the packets are reassembled in order and the message is then passed to the receiving user. The typical packet length is 1,024 bits (128 characters).

In the United States, ARPANET, the oldest of the packet-switching networks, now interconnects approximately 100 computers. This network is primarily intended for research-oriented users and is not licensed as a common carrier. TELENET provides local access to 85 cities in the U.S., Canada, and Mexico. Users in other U.S. cities can access the network via a dial-up service. Service to Puerto Rico, the United Kingdom, and Western Europe is available through gateway switching on a satellite network. Additionally, TYMNET was approved in early 1977 to operate as a public carrier. The TYMNET network contains over 70 nodes serving 63 cities. Plans are to increase this to 111 cities.

In addition, AT&T announced in September 1977 that it is proposing a packet-switching network to be called the Bell Data Network (BDN). The proposed system will provide communications processing functions and message distribution. The functions provided can be customized to the user's needs through a high-level programming language. The proposed network would connect hosts to hosts, hosts to terminals, or terminals to terminals. There are three priority levels for end-to-end transmission: priority 1 for 200 msec, priority 2 for 15 sec, and priority 3 for 30 min. The standard transmission time for non-priority data is a maximum of 4 hr. Data rates for the Bell network are asynchronous, operating at 300, 600, 1,200, and 2,000 bits/sec. Access to the network will be via a standard Bell 202 data set or its equivalent.

The three major types of services available from value-added networks are given in Table 9.1.1.3-1. The first type is datagram service, in which the host computer is responsible for control of the information. The responsibility of addressing, sequencing, and error correction rests entirely on the host. With the virtual-circuit network, the network appears to provide a dedicated line between the host and the destination. The majority of current networks provide this type of service. The third type is terminal emulation, in which the network assumes all responsibility for control and the host serves only as a receiving and sending device, much like a terminal.

One aspect important to value-added networks is that of network interfaces. The network interface takes on two major characteristics: first, the interface between the user and the network, and second, the interface between the network and a second network if internetwork communications are required. To aid in network interfaces, a common interface standard, specified as X.25, was adopted by the packet carriers. This standard specifies both physical and electrical characteristics of the user interface, in addition to requirements for the data structure. The standard also specifies the characteristics for the interface between

TABLE 9.1.1.3-1. TYPES OF SERVICE OFFERED BY VALUE-ADDED NETWORKS

TYPE OF SERVICE	COMPUTER RESPONSIBILITY	NETWORK RESPONSIBILITY
Datagram	Control, addressing, sequencing, error correction	Transfer of non-sequential data
Virtual Circuit*	Supply and receive sequential data	Appears as a dedicated line
Terminal Emulation	Supply and receive serial data (functions only as terminal)	Control, addressing, sequencing, error correction

*Most common type of service

packet-switching networks. This standard does not dictate the form used for transmission within the network, however. At the present time, the majority of packet networks provide interfaces compatible with X.25 and the others are phasing the standard into their systems.

Typical value-added networks operate with packet lengths of 1,024 bits, allowing 128 characters per packet. Low-speed interfaces are available for the small terminal user, with higher-speed lines of 1.2, 2.4, 4.8, and 9.6 kbits/sec available for the larger users. The typical network rate is 56 kbits/sec, but the actual data rate supported can run from 1/3 to 1/2 that, depending on network traffic codes and the number of users on the line.

9.1.1.4 Trends and Projected Developments in Terrestrial Networks - The trend in terrestrial communications networks clearly indicates a shift to an all-digital network. The existing analog networks will continue to be strong through the early 1980's, but digital equipment will be installed as the older analog equipment is phased out of service. Long-haul trunk lines will be the first part of the network to become all-digital, probably by 1985. The subscriber lines will continue to be dominated by analog circuits, although there is a shift to digital circuits for subscriber networks. This shift will occur first in rural areas and in urban data hookups. The digital systems offer noise reduction and cost decreases for long rural lines. Of course they offer the most straightforward approach for digital data transmission in urban areas. By 1985, the availability of subscriber digital circuits for use in voice/data applications should be good. Additionally, a method of data compression for digitized voice will be standard by this time, allowing voice to be sent at rates much below the 64-kbit/sec rates of the present systems.

The multiplexing speeds on high-traffic trunks will increase, allowing more data to be sent at faster rates. Several laboratory multiplexer systems are under development to extend the European system hierarchy to the fifth level at 565.148 Mbits/sec. One design under investigation uses a pure binary chain and subnanosecond ECL circuits. This system uses a 2.6- by 9.5-mm coaxial transmission line and requires repeaters at intervals of 850 m. Another experimental laboratory multiplexer system achieves the 565.148-Mbit/sec rate using a four-level digital signal. This system uses ECL operating near 0.5-nsec speeds. The system was tested over an 1,800 m cable with a measured noise margin of 6.3 dB. Multiplexers for raising the digital hierarchy of other countries are being developed and it is expected that rates approaching 1-Gbits/sec will be in use by the 1985 timeframe.

The use of optical fibers for network trunk lines will be quite common in the 1980-1985 timeframe. Experimental networks tested by AT&T,

GTE, and several telephone companies in Europe and Japan have proved that optical fiber links are both reliable and practical. These experimental links range from low-rate (192-kbit/sec) and high-rate (44.7-Mbit/sec) in-service prototype links to higher-rate (140-Mbit/sec) nontraffic test links. Results have been encouraging for all of the major tests conducted thus far. These systems are comparably priced with coaxial systems at present, and the optical systems are expected to offer a price advantage as the cost of the fibers is lowered and as associated equipment becomes more reliable. The use of fiber optics for other systems will also increase in use. Fiber-optic links between computers have already been installed by several banks and their use for other wideband links will undoubtedly increase, especially for applications that are either space- or weight-limited. There have been proposals for using optical fibers to reduce the weight of ships, airplanes, and even spacecraft. In terrestrial networks, fiber-optics have the advantage of fitting into existing communications ducting that would otherwise be too crowded for additional coaxial lines. Fiber-optics will be used commonly by 1980 for the transmission of digital data over short links not requiring repeaters. By 1985, fiber-optic systems will be used for prototype long-haul high-rate communications.

Value-added networks are expected to mature as interface standards provide for the orderly interconnection of international networks. The number of cities receiving services of this type will also increase to accommodate the increasing number of machines requiring data transmission service. Small, inner-city packet-switching networks will develop as banks and other large data users implement their own networks. Additionally, value-added networks will be implemented by some of the major common carriers, as outlined by Bell's proposal for the BDN. However, the types of packet-switching services offered will likely be determined more by regulations and restrictions than by the technical limitations on equipment.

The amount of data that can be transmitted over voice-grade analog lines is expected to increase during the 1980-1985 timeframe.

However, this will occur more as an improvement in modem technology than as a change in network quality. Band rates of 9.6 kbits/sec over a multipoint line should be achievable in the 1980-1985 timeframe. An increase to 14.4 kbits/sec over a point-to-point line could be realized by 1980, although an increase much above this will be unlikely due to phase nonlinearity and noise limitations. Harris Radiation Laboratories are currently working under contract to develop a 16-kbit/sec modem for the Rome Air Development Center. Specifications for this modem are presently unavailable and details of an implementation schedule are not known.

Terrestrial networks in general will experience growth and increased capability. Many of the networks will become more dependent on satellite links, for both domestic and international traffic. Additionally the networks will provide more wideband services to their customers.

9.1.2 Satellite Networks

This subsection addresses satellite network technology from the viewpoint of the availability and characteristics of satellite networks for telecommunications. Emphasis is given to the areas of wideband digital data transfer and the development of the "small Earth terminal" for individual user accessibility. Other aspects of satellite technology, such as orbiting spacecraft-to-ground communications and hardware components (e.g., transmitters, receivers, and antennas), are presented in Section 5.

9.1.2.1 State of the Art in Satellite Networks - Satellite telecommunications networks have been in existence since 1965, when INTELSAT I was placed in orbit to provide communications coverage between the United States and Western Europe. Both the number of communications satellites and the number of satellite networks have grown rapidly since that time, with a major growth having taken place since the early 1970's. The number of satellites remains less than the number of satellite networks, however, since a number of networks (both in the U.S. and abroad) lease capacity on satellite channels in lieu of operating their own satellites.

Table 9.1.2.1-1 presents a listing of the regional (multi-country coverage) satellite programs that are either in operation or are planned for operation in the 1980-1985 timeframe. Listed in this table are the program or sponsor, the classification according to use, the status of the program, the date of system operation or planning, and the frequencies used by the systems. Table 9.1.2.1-2 presents the same type of information for domestic (single-country coverage) satellite network programs. Networks using the transponders of a leased satellite are included in the table with the facility name indicated under the frequency column. Table 9.1.2.1-3 lists the satellite network system characteristics for a group of communications satellites. Characteristics include the modulation and multiple-access technique, the diameter of the required Earth station antenna, and the capacity of the satellite. The key acronyms for the modulation and multiple-access techniques are presented in Table 9.1.2.1-4. The U.S. domestic satellite systems are given in Table 9.1.2.1-5, including the company, Earth station locations, services offered, satellite capacity, and the program status.

Several foreign agencies have major programs underway to develop advanced communications satellite networks. The European Communications Satellite (ECS) system is being developed with a goal of having over 20 Earth stations linked together by the early 1980's. The Japanese have a program underway with three communications satellites to be launched in 1978. Near-term goals of the Japanese efforts are: 1) to broadcast

TABLE 9.1.2.1-1. REGIONAL SATELLITE NETWORK PROGRAMS

PROGRAM OR SPONSOR	CLASS OF USE						STATUS				DATE OF OPERATION OR PLANNING	FREQUENCY (GHz)	
	FIXED	MOBILE	BROADCAST	EXPERIMENTAL	MILITARY	PROPOSED	OPERATIONAL	UNDER CONSTRUCTION	IN PLANNING	UNDER CONSTRUCTION		UP	DOWN
INTELSAT	X						X	X			1965	6	4
USSR	X						X				1965	6	4,1
European Space Agency	X							X			1977	14	11
Norway/North Sea*	X						X				1975	via INTELSAT	
Arab System	X									X	1980-1990	C	2.5
COMSAT G. (MARISAT)		X					X				1975	UHF and L**	
												6	4†
ESA/COMSAT G. (AEROSAT)		X							X		1979	VHF and L**	
													C†
ESA (MAROTS)		X					X				1977		L**
												14	11†
IMCO (INMARSAT)		X								X	1980+		L**
												C or Ku†	
European Broadcasting Union			X							X	1980+	Ku	
Symphonie				X			X				1975	6	4
CTS				X			X				1975	14	11
DSCS	X	X			X		X	X			1966	8	7
PLTSATCOM	X	X			X			X			1977	UHF**	
													8†
LES Series	X	X		X	X		X				1965	UHF and K	
Andean Nations*	X					X				X	?	via INTELSAT	

*Earth stations only
**Satellite to mobile terminal
†Satellite to fixed ground terminal

TABLE 9.1.2.1-2. DOMESTIC SATELLITE NETWORK PROGRAMS

PROGRAM OR SPONSOR	CLASS OF USE						STATUS				DATE OF OPERATION OR PLANNING	FREQUENCY (GHz)	
	FIXED	MOBILE	BROADCAST	EXPERIMENTAL	MILITARY	PROPOSED	OPERATIONAL	UNDER CONSTRUCTION	IN PLANNING	UNDER CONSTRUCTION		UP	DOWN
TELESAT - Canada	X						X				1973	6	4
Western Union	X						X	X			1974	6	4
RCA SATCOM	X						X	X			1974	6	4
American Satellite Corporation*	X						X				1974	via WESTAR	
COMSAT/ATT	X						X				1976	6	4
Algeria*	X						X				1975	via INTELSAT	
Indonesia	X						X				1976	6	4
SBS	X							X			1981	14	12
Philippines*	X									X	1976	6	4
India	X		X							X	?	C or Ku	2.5
FRG			X							X	=1985	Ku	
ATS-6				X			X				1974	K,C,S,UHF, VHF	
Japan-Communication	X			X				X			1977	6	4
Japan-Broadcast			X	X				X			1978	14	12
SIRIO				X				X			1977	17	11
Argentina*	X					X				X	?	via INTELSAT	
Australia	X					X				X	1980's	C,Ku,S	
Brazil	X		?		?	X			X		1978	6	4
Denmark*	X					X				X	?	via INTELSAT	
Iran	X		?		?	X				X	?	--	
Malaysia*	X					X				X	1975	via INTELSAT	
United Kingdom						X					-	--	
Asian Nations	X					X				X	?	--	
Systems Using INTELSAT*	X	X				X	X	X		X	Various	via INTELSAT	

*Earth stations only.

TABLE 9.1.2.1-3. SATELLITE NETWORK SYSTEM CHARACTERISTICS (Ref. 9-3)

SYSTEM	MODULATION AND MULTIPLE ACCESS	EARTH STATION ANTENNA DIAMETER (m)	CAPACITY PER SATELLITE
INTELSAT IV INTELSAT IV-A INTELSAT V	FM/Video FDM/FM/FDMA SPADE	29.5 for STD. A 10 for STD. B	7,500 Channels + TV + SPADE 12,000 Channels + TV + SPADE 27 Channels of 36 MHz
USSR-Molnia	FM	12 25	One TV Channel +Unspecified Telephones
TELSAT-Canada	Single-Carrier FM Multi-Carrier FM Single-Channel/Carrier Delta Modulation PSK/TDMA	Heavy Route - 30 Network TV - 10.1 Northern Telecommunications - 10.1 Remote TV 8.1/4.7 Thin Route 8.1/4.7	12 Transponders of 36-MHz Bandwidth
Western Union (WESTAR)	SSB/FM/FDM Single- and Multiple-Access Video, SCPC, PSK/TDM/TDMA	10	12 Video Channels (One-Way) or 14,400 FDM Voice Channels (One-Way)
RCA SATCOM	FDMA: FDM/FM and PCM/PSK for Voice Data, 4 PSK for Digital Data, FM for Monochrome or Color TV TDMA: PCM/PSK for Voice/Data	13 10	24 Video Channels with 34-MHz Bandwidth 9,000 Channels per Transponder
COMSAT/ATT-GTE (COMSTAR)	FDM/FM Digital Transmission, 4PSK	30 13 10	28,800 One-Way Telephone Channels or 1,073-Mbit/sec Data
European Space Agency (OTS/ECS)	4-Phase PSK FM Video TDMA	European A 13 + Spot Beam European B 3	One 120-MHz Transponder One 40-MHz Transponder One 5-MHz Transponder
Indonesia	FDM/FM Multiple Carriers per Transponder SCPC/FM Demand Assigned	9.8 7.3 4.0	12 Transponders of 36-MHz Bandwidth

TABLE 9.1.2.1-3 - Continued

SYSTEM	MODULATION AND MULTIPLE ACCESS	EARTH STATION ANTENNA DIAMETER (m)	CAPACITY PER SATELLITE
COMSAT General (MARISAT)	Voice: SCPC-FM Data: 2-Phase Coherent PSK	1.22 Mobile Terminals 12.8 Shore Terminals	9 Voice Channels (Both Ways) 110 Teleprinter Channels (Both-Ways)
CTS	FM Video FM Sound Broadcast 10 Channels (FDM) of FM Duplex Voice	Ten 0.91 Eight 2.43 Two 3.05 Two 9.14	1 TV Channel 1 Sound Broadcast Channel 10 Duplex Voice Channels
ESA/COMSAT General (AEROSAT)	Voice: NBFM, PDM and VDSM Data: PSK/FSK	Low-Gain Airborne Antennas	Five 80-kHz Communications Channels for Ground-to-Air and Surveillance Fifteen 40-kHz Communications Channels for Air-to-Ground Two 80-kHz Communications Channels for Ground-to-Ground One 400-kHz or 10-MHz Experimental Channels
ESA (MAROTS)	FDM TDM FDMA TDMA	0.6 for Mobile Terminals	Shore-to-Ship: Up to 50 Voice/High-Speed Data Channels Ship-to-Shore: Up to 60 Voice/High-Speed Data Channels Shore-to-Shore: Up to 3 Voice/High-Speed Data Channels
Satellite Business Systems (SBS) (1981)	TDMA 4-PSK	5 7	10 Transponders at 36-MHz Bandwidth (328 Mbits/sec)
Symphonie	Analog and Digital	16 12 8 4	1,200 One-Way Telephone or 2 Color TV Channels
ATS-6	FM, Video	25.9 3 (Various)	2 Video Channels at 2.6 GHz or 1 Video Channel at UHF (860 MHz) C-Band Transponder has 40-MHz Bandwidth 1.5-GHz Transponder has 12-MHz Bandwidth

TABLE 9.1.2.1-3 - Concluded

SYSTEM	MODULATION AND MULTIPLE ACCESS	EARTH STATION ANTENNA DIAMETER (m)	CAPACITY PER SATELLITE
SIRIO	PCM-PSK, 2-Phase for Voice (Narrow-Band Communications) FM or Digital for TV (Wideband Communications)	14.5 for Stations in Italy 14 for Stations in Finland 12 for Stations in U.S. Various Smaller sizes Down to 1.2 for Shipboard Terminal	Twelve 100-kHz Telephone Channels 1.5 MHz Total Bandwidth or one 4-MHz Baseband or TV
DSCS II	Stage Ia of Program: FDMA and CDMA Stage Ib of Program: FDMA and CDMA Stage Ic of Program: FDMA and CDMA, Phasing into TDM/PCM Stage 2 of Program: TDMA	18.2 for Fixed Terminals 0.8 for Airborne Terminals	1,300 Duplex Voice Channels or 100-Mbit/sec Data Total of 410 MHz of Transponder Bandwidth
SKYNET	CDMA in 20 MHz Channel FDMA in 2-MHz Channel	I: 12.8 II: 12.2 III and IV: 6.4 V: 1.8 "Scot": 1.1	One 20-MHz Channel One 2-MHz Channel 24 (2,400 bit/sec) Data Channels or 280 Voice Channels
NATO	FDMA/FDM (Clear Mode) CDMA (Jamming Mode)	12.8	
FLTSATCOM			9 UHF and 1 SHF Uplink 10 UHF Downlink Each UHF has 25-kHz Bandwidth
LES Series	DPSK Downlink QPSK Conferencing Link 8-ary FSK Foward Uplink 8-ary MFSK, Hopped at 200/sec	1.2 for ABNCP Terminal (Lincoln Labs) 0.92 for Airborne Terminal AN/ASC-22 4.6 for Navy Terminal	36 to 38 GHz: 10 kbits/sec, DPSK to other LES Satellites 20 kbits/sec, DPSK to ABNCP Terminal 8-ary FSK Forward Uplink; QPSK Conferencing Uplink 50 k-ary Symbols/sec from ABNCP Terminal 75 bits/sec to Navy Shipboard Terminal UHF: 50 8-ary Symbols/sec to Aircraft 100 8-ary Symbols/sec from Aircraft

TABLE 9.1.2.1-4. ACRONYMS FOR MODULATION AND MULTIPLE ACCESS

DPSK	Differential Phase Shift Keyed
QPSK	Quadriphase Shift Keyed
FDMA	Frequency Division Multiple Access
TDMA	Time Division Multiple Access
SCPC	Single Channel per Carrier
SPADE	Single-Channel-per-Carrier PCM Multiple Access Demand-Assigned Equipment
NBFM	Narrow-Band Frequency Modulation
CDMA	Code Division Multiple Access
MFSK	Multiple Frequency Shift Keyed
VSDM	Variable Slope Delta Modulation

...DUCIBILITY OF THE
...AL PAGE IS POOR

TABLE 9.1.2.1-5. U.S. DOMESTIC SATELLITE SYSTEMS

COMPANY	EARTH STATION LOCATIONS	SERVICES	SATELLITES	STATUS
Western Union	New York Chicago Los Angeles Dallas Atlanta Honolulu	Data, voice, and video leased private line	Two 12-channel HS-333s (WESTAR) Two advanced WESTARS	Operational July 1974 Operational 1980
American Satellite Corporation (A.S.C.)	New York Los Angeles Dallas Fairchild AFB, WA Loring AFB, ME Centerville Beach, CA Moffett Field, CA Offut AFB, NB Monterrey, CA Chicopee, MA Orlando, FL San Francisco	Data, voice, and video leased private line	Phase 1: lease of 3 channels in WESTAR Phase 2: 12-channel HS-333s	Operational July 1974
RCA Global Communications RCA Alaska Communications	New York Los Angeles San Francisco Juneau, AK Honolulu Washington, DC San Juan, PR Houston Prudhoe Bay, AK Cordova, AK Nome, AK Valdez, AK Talkeetna, AK Bethel, AK Yakutat, AK	Data, voice, and video leased private line plus MTT within Alaska and between Alaska and CONUS	Phase 1: lease of 2 channels in ANIK (or WESTAR) Phase 2: two 24-channel satellites built by RCA Astro-Electronics	Operational January 1974
AT&T and GTE Satellite	New York Chicago San Francisco Atlanta Los Angeles Tampa Honolulu	MTT (no private line except to the U.S. Government for 3 years)	Leased from COMSAT General by AT&T	Earth station authorizations granted
COMSAT General	TT&C only (Santa Paula, CA, and Southbury, CT)	Lease of transponders to AT&T	Three 24-channel satellites (modification of INTELSAT IV) built by Hughes Aircraft Company	Leased to AT&T
General Electric Company	Valley Forge, PA Daytona Beach, FL	Company communications equipment and system development	Transponders leased from Western Union	Operational
Satellite Business Systems (SBS)	TT&C-New Jersey, CA User terminal - on location	Digital data, voice, and facsimile	Two 10-channel satellites, built by Hughes Aircraft Company	To be operational January 1981

television from space on a national coverage basis, 2) to establish a wideband digital communication network using Time Division Multiple Access (TDMA) and operating in both the 4- to 6-GHz and the 20- to 30-GHz bands, and 3) to conduct extensive testing of communications in the 30- to 35-GHz frequency range. Other long-term programs to make Japan a communications satellite leader are in the planning stages. Additionally, several other national and regional networks are planned abroad, but these will have only marginal impact on the world communication system.

The use of communication satellites has been rapidly increasing since 1975 and many new uses are proposed. Several of the U.S. domestic satellite carriers have transponders leased for use by cable television companies. The decision of the Federal Communications Commission in late 1976 to approve the use of 4.5-m antennas for receive-only terminals for cable companies has accelerated the use of satellites for distribution of "pay television". These 4.5-m stations can be built and maintained at lower cost than conventional 10-m stations. Estimates are as low as $18K per station in quantity. Additionally, the Corporation for Public Broadcasting (PBS) is currently in the process of changing from a terrestrial to a satellite network for distribution of educational television. The network, to be fully operational by late 1978, will consist of 163 Earth stations, four of which will have uplink capability. The network will use leased satellite channels operating at C-band. Other systems, operating at narrow bandwidths, have been proposed for the distribution of music, both for radio networks and for consumer background music. Hughes recently introduced a station with a 3-m antenna for reception of two-channel audio signals of 15-kHz bandwidth. The cost of the system is $5K, and it is expected to offer an alternative to leased line audio distribution.

The use of broadcast satellites on an in-depth experimental basis began with the launch of the Applications Technology Satellite (ATS). Since then, several other satellites have incorporated broadcast distribution on an experimental basis, including the Communications Technology Satellite (CTS). Ground stations for these experimental

broadcast satellites use antennas as small as 1.0 m, with 1.6 m being common for the 12-GHz broadcast band. The Japanese are planning a broadcast satellite for 1978 that will also operate in the Ku-band. The ground stations for this satellite will range from 2.5 to 4.5 m. As the power of satellite transmitters increases, the use of broadcast satellites to transmit to small receive-only stations will increase.

In the area of wideband data transfer by communications satellite, the current state of the art is probably best presented by the proposed Satellite Business Systems (SBS) network. SBS, to be operational in early 1981, will be the first totally digital data relay system operating at Ku-band. Each subscriber will have an Earth terminal located on the premises, and will have access to the network without the use of landlines. The system will be an all-digital link with capabilities for data, voice, facsimile, and freeze-frame video. The unattended Earth terminals for the SBS will use either a 5- or 7-m antenna and have a 500-W power amplifier for uplink. These ground stations feature a time-division multiple-access burst modem with rate capacity of up to 43 Mbits/sec and options for 48-Mbit/sec operation. The user's throughput data rate can be controlled from the SBS control center on a need basis. The achievable bit error rate for the system is 10^{-5} for 99.5% of the time. An optional 10^{-7} BER will be provided through error correction coding. The first phase of testing was conducted in early 1977 using leased transponders at C-band. Phase 2 tests are scheduled for late 1977 and early 1978 using the Ku-band facilities of the CTS. These tests will be conducted with the participation of several large corporations.

The use of small terminals for reception of data from communication satellites is currently feasible. With the higher satellite powers and the lower noise receivers of today, the use of small antennas in a receive-only mode presents no problems. For interactive terminals, however, the transmit mode can present several problems. The receiver noise temperature for spaceborne transponders runs several decibels above those for uncooled ground stations. In addition, the smaller-diameter antenna on the ground has a lower gain, requiring more transmitter power on the ground. While current technology does not present

a limitation on available power, the beamwidth of the small terminal antenna is wide and the sidelobe level is high. Much of this power is beamed into places where it is not only wasted but where it may interfere with other satellite systems. Geosynchronous communications satellites are already working in an interference environment, and the additional interference from small terminal antennas cannot be tolerated. The Federal Communications Commission currently limits the uplink antenna size to a minimum of 10 m at C-band. Smaller antennas (such as the SBS 5- and 7-m) may be used at higher frequencies because the beamwidth is dependent on frequency. These small terminals are of course less expensive than the larger 10- to 30-m manned terminals, but they are still quite expensive, with costs estimated from $350K to $500K per interactive terminal.

One application in which the receive-only small terminal would be favorable is for multipoint distribution of data. The originating ground station would relay the data through the satellite where a number of small terminal stations could receive them on site and process the data to extract those portions of interest to them.

Typical performance characteristics for satellites operating in the more common communications bands are presented in Table 9.1.2.1-6. Channel bandwidth is limited to 500 MHz for both C- and Ku-bands; but in the 20- to 30-GHz band, bandwidths of up to 1,500 MHz are allocated. For each individual transponder, 36 MHz is typical for C-band operation. At Ku-band, 85-MHz bandwidth per transponder is common, with a few special exceptions using up to 250 MHz. Typical transponder bandwidths have not been established at 20 GHz, but bandwidths of a few hundred megahertz are technically feasible at these frequencies. The general error performance at C-band is slightly better than at the higher frequencies, but extreme spectral crowding exists at C-band, leaving little room for additional services. At both Ku- and Ka-bands, atmospheric conditions exhibit a great effect on the error performance. Severe attenuation due to rain limits the usefulness during extreme weather conditions. Additionally, degradation at Ka-band occurs from water

TABLE 9.1.2.1-6. PERFORMANCE CHARACTERISTICS FOR COMMUNICATIONS SATELLITE BANDS

CHARACTERISTIC	C-BAND	Ku-BAND	Ka-BAND
Frequency	4/6 GHz	11/14 GHz	20/30 GHz
Channel Bandwidth	500 MHz	500 MHz	1,500 MHz
Transponder Bandwidth	36 MHz	85 MHz	-
Error Performance	10^{-6} BER	10^{-5} BER	10^{-4} to 10^{-5} BER
Special Considerations	Extreme spectral crowding	Attenuation from weather	Extreme attenuation from weather
Suitability for Small Terminals	Receive only	Good for interactive	Good for interactive

vapor in the atmosphere. For users of small antenna systems, the higher bands have some advantages. At C-band, small terminal use is limited to receive-only applications. At Ku-band, an interactive terminal becomes feasible for 4- to 5-m antennas. The equivalent antenna for an interactive small terminal operating at 20 to 30 GHz is only about 2 m. Component technology can currently support the equipment necessary for operation at these higher frequencies, but there are currently no commercial satellites supplying services in this band. Therefore, most of the satellite users will continue for the time being to interface with the network by land lines rather than by their own interactive terminals.

9.1.2.2 Trends in Satellite Networks - The general trend in data systems is to transmit more data in less time. Future communications satellites will increase in capacity as technology provides hardware with advanced capabilities. Frequency reuse will become more common as a scheme to double the transmission capacity of any one frequency by using orthogonal polarization to keep the two data channels separate. This frequency reuse scheme is currently being used at the 4- to 6-GHz band and will be used at both the 4- to 6-GHz and the 11- to 14-GHz bands with the implementation of INTELSAT V in 1979.

The 11- to 14-GHz band will be used to provide more channels and greater bandwidth. As technology advances, the 20- to 30-GHz band will be used. The use of these higher frequency bands will result in narrower antenna beamwidths and higher gains and will produce a decrease in ground antenna size. Additionally, the satellite transmitter power will increase at both the 4- to 6-GHz and the 11- to 14-GHz bands. This power increase will be made possible by better TWTs, increased prime power collection, and more directive antenna patterns. As with all systems, cost is a major factor. The launch costs of future satellites launched from the Space Shuttle will be an estimated 40% less than present costs. This will allow more of the system budget to be spent on the spacecraft and will allow increased weight to achieve increased capability.

The ground stations in the satellite networks will become smaller. The higher frequencies and higher power of ground stations, along with the advancement of better antenna designs, will allow use of small terminals with a minimal amount of interference to other satellite systems. Depending on the type of signal, the ground station antenna will range from 10 m for high-rate digital data down to 0.6 m for TV reception from high-power satellites. The ground station receivers will also be improved by lower noise amplifiers and solid-state controllers.

Communications satellite systems in general will continue to increase in capability. The existing systems will be updated in both capacity and capability (e.g., INTELSAT V and Western Union's TDRSS/Advanced WESTAR). New designs will provide for increased data rates,

both on a per-user and a per-satellite basis. The total available bandwidth will be increased through frequency reuse, combination of C- and Ku-band operation, and the migration to the 20- to 30-GHz band, where bandwidths of up to 1,500-MHz have been allocated. Additionally, the number of networks will increase as more companies start networks of their own. The trend for these networks will be for all-digital capability with ground stations located at or near the user. It has been predicted (Ref. 9-4) that 84 telecommunications satellites will be launched by the Western World between 1980 and 1990. Of these, 29 will be for the United States, 13 for Europe, and 16 for international organizations.

Figure 9.1.2.2-1 presents the trend in the number of U.S. domestic communications satellites available between 1975 and 1990. In addition to an increasing number of satellites, the number of transponder channels per satellite will also increase. Figure 9.1.2.2-2 gives the trend in the number of wideband transponder channels available for domestic communications. The cost of a wideband transponder channel will decrease as more channels become available, with the projected cost being $1M per year per transponder by 1990, as shown in Figure 9.1.2.2-3.

The number of users will increase as more channel capacity becomes available. Many of the users will have narrow bandwidth needs (such as real-time data connection between company teletypes) and will connect to the satellite network by means of the terrestrial network. Other users will have a need for wideband data links and will interface with the network through small terminals located at or near their premises. The flexibility available to these users will be increased by the use of real-time reassignment of a time-division multiple-access system. This will allow the throughput rate to be allocated to the user as needed.

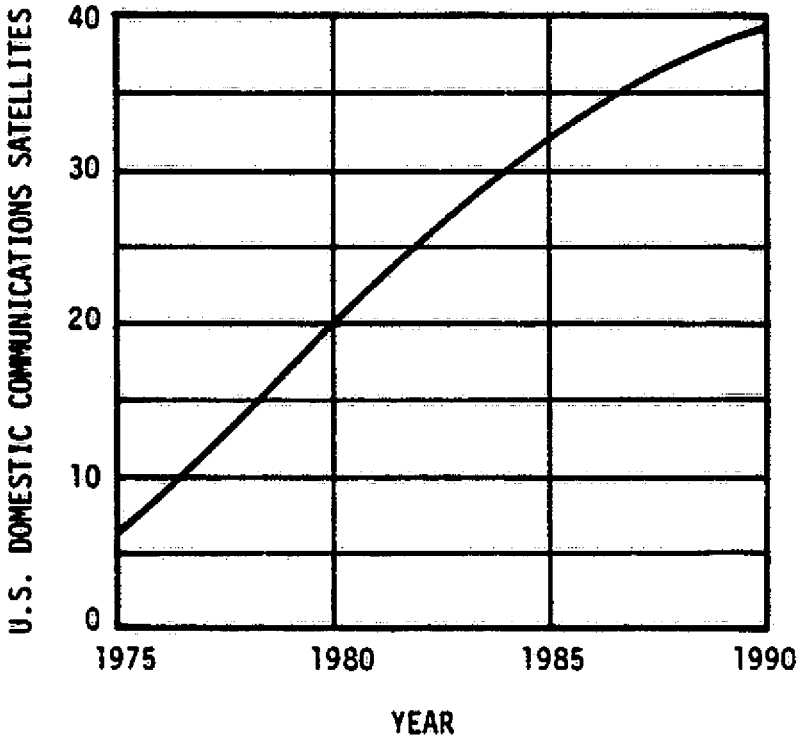


FIGURE 9.1.2.2-1. TRENDS IN THE NUMBER OF U.S. DOMESTIC COMMUNICATIONS SATELLITES

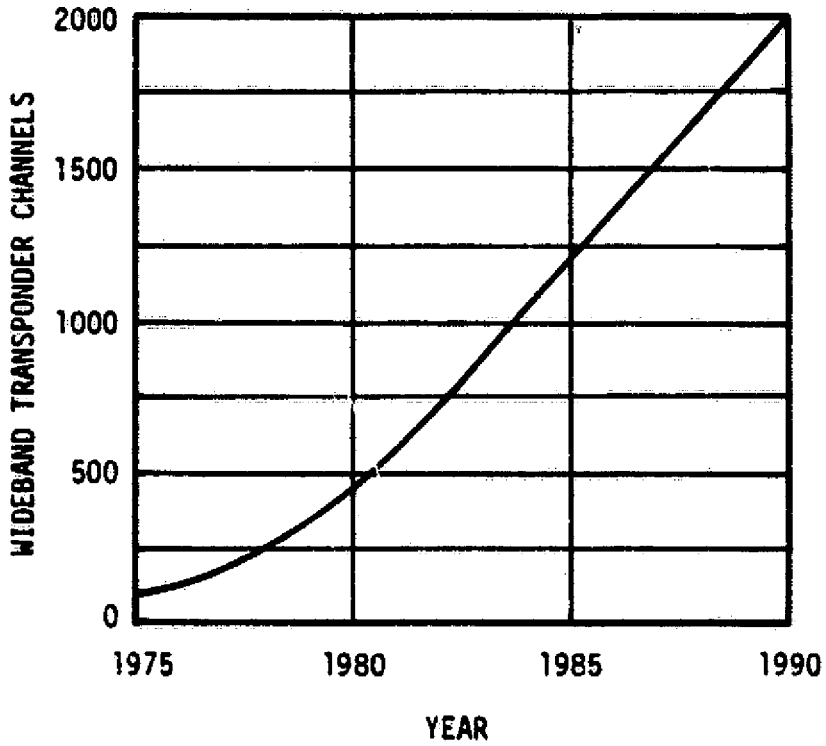


FIGURE 9.1.2.2-2. TRENDS IN THE NUMBER OF WIDEBAND TRANSPONDER CHANNELS FOR DOMESTIC COMMUNICATIONS

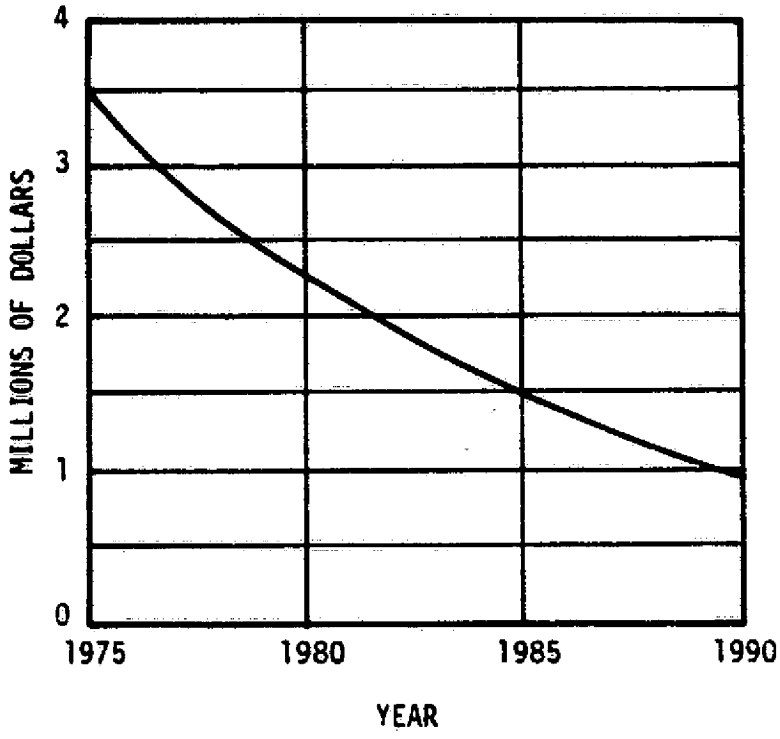


FIGURE 9.1.2.2-3. TRENDS IN THE ANNUAL COST OF A LEASED WIDEBAND TRANSPONDER CHANNEL

9.1.2.3 Projected Developments in Satellite Networks - While the 4- to 6-GHz band will continue to be used, the 11- to 14-GHz band will become the predominant band for wideband data and ground stations employing small-diameter antennas. At Ku-band, RF bandwidths of 85-MHz (versus 36-MHz at C-band) will be common, with a few special systems having links of up to 250-MHz bandwidth. The majority of these satellite networks will relay digital data, with the most common exception being analog FM links for broadcast satellite television.

The ground stations for the 11- to 14-GHz band will use mostly 5- to 10-m antennas. Receive-only stations for TV will use 1- to 2-m antennas at Ku-band, while 4.5-m antennas will continue to be used for receive-only TV operating at C-band. Interactive terminals will operate at Ku-band or above, making the use of small terminals feasible. These small terminals can be configured for transmission of wideband data at these higher frequencies. The cost of a wideband channel is relatively high, but the per-user cost can be kept low if transmissions are scheduled ahead of time and if multiple users receive the data in a broadcast mode.

The number of satellites and satellite networks will increase as more communications data shifts from terrestrial to satellite networks. The transfer of both high-speed data and commercial television will be accomplished by satellite networks instead of ground distribution networks. These networks will have the ground station at the user and will reduce the long-term cost of the system when compared with wideband line network charges. Narrowband systems will also use satellite links, especially for the distribution of a single signal over a wide area (as in radio news networks or background music suppliers).

The hardware technology to implement these satellite communication systems (see Section 5) either currently exists or will exist during the 1980-1985 timeframe. Care must be taken, however, to make optimum use of the satellites and satellite networks. A recent study by the Office of Telecommunications (Ref. 9-5) shows that a direct satellite communications system has a total capacity of 2,000-Mbits/sec

from each satellite. By comparison, the AT&T COMSTAR series has a capacity of 1,073 Mbits/sec and the SBS a capacity of only 328 Mbits/sec. As the geosynchronous orbits continue to be filled, it becomes increasingly important for each satellite to produce the greatest benefits possible. Multifunction satellites (such as the TDRSS/Advanced WESTAR) increase the utilization of orbital space by providing several services in one satellite. This trend will become more common as multifunction satellites provide broadcast, point-to-point data links, and multipoint data relay, all from the same satellite. This trend will reduce the amount of orbital space required as well as reducing the spacecraft cost (as compared with using two or three separate satellites). This technique will become feasible with the operation of the Space Shuttle, allowing capacity for heavier systems and the ability for satellite retrieval in the event of system failure.

Another technique that will allow a more optimum use of satellite capacity is a scanning spot-beam system, such as that proposed by Bell Laboratories. Their proposal utilizes fixed spot beams for coverage of major cities, with a phased-array antenna scanning the remainder of the country. Sweep rate for the system is 0.01 sec and stations are polled by a time-division multiple-access system, much like the SBS scheme. The system will allow a more efficient use of the radio spectrum since the beam is limited to 1% of the nation at any instant as opposed to a beam covering the entire country. This allows frequency reuse techniques that increase satellite capacity by up to three times while actually lowering the required EIRP by as much as 20 dB. The system uses a 10-ft-diameter Earth-station antenna located near the user. Bell is currently designing a prototype of the system, but has no comment on when it may be implemented.

9.2 TELECOMMUNICATION HARDWARE

Telecommunication hardware includes all the terminal and interface equipment that makes up a communications network except for the long lines or radio links. These include terminals, data multiplexers, modems, front-end processors, and network processors. This section covers these pieces of hardware with the exception of terminals, which are covered in Section 10.

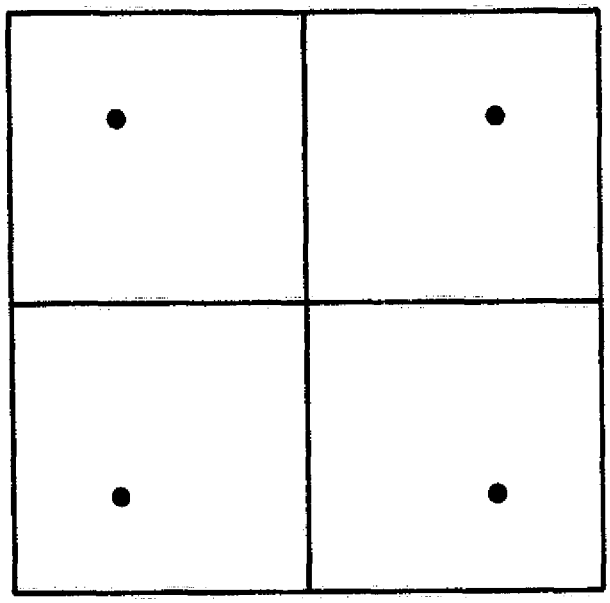
Telecommunication hardware is changing as the requirements of the data network change. These changes are generally toward increasing capability, not only to better perform the traditional function of the device but to increase its role in the total communication system. Many of these devices are acquiring a processing capability, allowing them to perform functions that were once either not performed or done at the central computer facility. This is leading to a distributed network with capacity spread around the network rather than concentrated at one point.

9.2.1 Modems

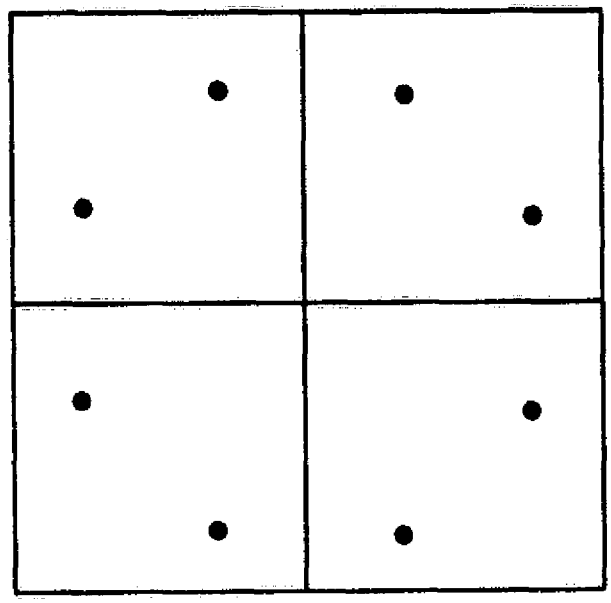
Modems (short for modulator/demodulator) are the interface devices used to connect digital equipment to analog transmission channels. The modem provides the transition of serial digital data from a baseband pulse train to a modulated carrier compatible with the 300- to 3,300-Hz bandwidth of standard voice grade communications channels. The modem uses multilevel modulation techniques and complex equalization schemes to provide bit rates that exceed the normal channel capacity of the line. Advances in modem technology account for a major improvement in the effective throughput of data over telecommunication channels.

9.2.1.1 State of the Art in Modems - Modems operate either synchronously or asynchronously, depending on whether the receiving and transmitting terminals operate in phase. Asynchronous modems are generally limited to data rates below 1,200 bits/sec, as might be required for a teletype terminal. Asynchronous modems may be either acoustically coupled to a standard telephone handset or hardwired to the communications line. Frequency-Shift-Keyed (FSK) modulation is normally used for asynchronous modems. As the data rate increases above 2-kbits/sec, multilevel modulation is required to maintain the signal within the limited line bandwidth. These modulation techniques require the modem to be synchronized with the sending terminal. Synchronous modems normally operate between 2.4 and 9.6 kbits/sec. Several modulation techniques are used to achieve the high data-rate-to-bandwidth ratio required for these types of rates. Figure 9.2.1.1-1 illustrates the signal constellation patterns for several of the more popular methods. The constellation at (a) shows the pattern for four-phase DPSK used for sending 2,400 bits/sec. An eight-phase DPSK (b) is the standard for 4,800 bits/sec. Several standards exist for the 9,600-bit/sec rate. The U.S. common carriers use a 4×4 Quadrature Amplitude Modulation (QAM) technique, as shown in (c). The independent modem manufacturers use a two-level, eight-phase PSK, as illustrated in (d). No matter which method is used for 9,600-bit/sec modulation, the modem must handle 16 signal conditions. The number of conditions sets a practical limit on the amount of data sent over a limited bandwidth line. To achieve even 14.4 kbits/sec, it is necessary to maintain 24 conditions and as the number of conditions goes up, the individual points become more difficult to decode. The phase shift and noise on the line become extremely critical and reliability goes down unless extensive coding techniques are employed. Some multiplexed modems accommodate a rate of 19.2 kbits/sec, but this is achieved by sending data at a rate of 9.6 kbits/sec over two voice-grade lines.

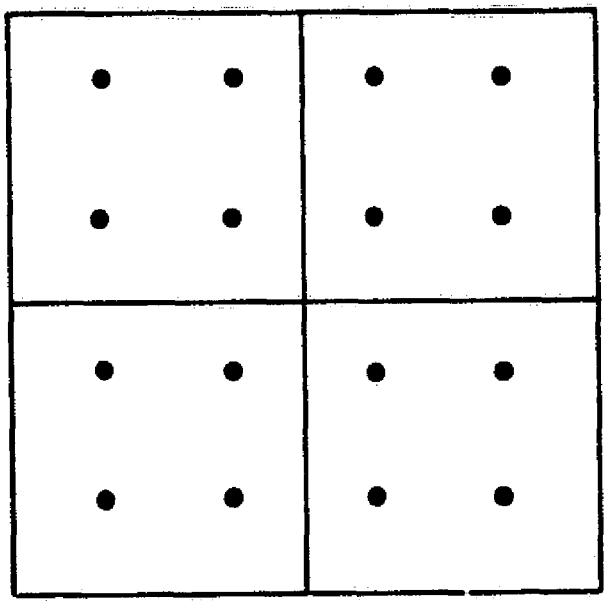
Modems have reached a high level of sophistication compared with the early units of the 1960's. Some current designs employ microprocessors for control. Others use MSI or LSI for all functions. Many of the present



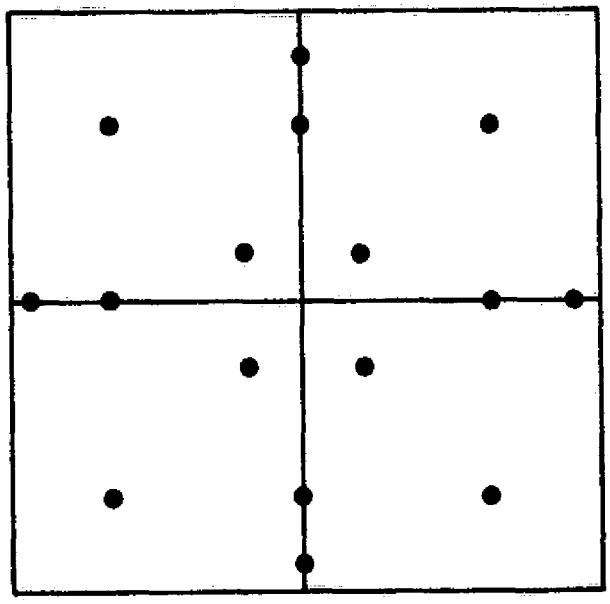
(a) FOUR-PHASE DPSK STANDARD FOR 2,400 bits/sec



(b) EIGHT-PHASE DPSK STANDARD FOR 4,800 bits/sec



(c) 4×4 QAM U.S. CARRIER STANDARD FOR 9,600 bits/sec



(d) TWO-LEVEL, EIGHT-PHASE PSK INDEPENDENT MODEM MANUFACTURERS' STANDARD FOR 9,600 bits/sec

FIGURE 9.2.1.1-1. SIGNAL CONSTELLATIONS FOR SEVERAL MODEM MODULATION TECHNIQUES

modems have status lights that monitor in the range of 9 to 11 vital interface conditions. In the event of a system failure, a cursory examination of the modem interface conditions will isolate the problem. Some modem models offer automatically adaptive equalizers. These units determine the equalization required for the line in use and then set an active filter system to achieve this value. In this way, a dial-up line can be used without the modem operator having to adjust the modem equalization. These modems, when used with dial-up lines, are capable of achieving data rates once possible only with leased lines. The cost of these modems is quite high, however, and their implementation cannot be justified by some users. Likewise, modems are available with forward error corrective coding techniques, but the cost is too great for the typical user.

A listing of the characteristics for some typical state-of-the-art modems is given in Table 9.2.1.1-1. All of the modems listed are for rates at or below 9.6 kbits/sec except the Codex Model 8310. This is a high-rate modem capable of sending up to 64 kbits/sec over a wideband line. In most cases, the modems can operate over either a switched or point-to-point line at rates of 4.8 kbits/sec or less, but a conditioned point-to-point line is required for data rates above 4.8 kbits/sec.

Since modems provide the link between a variety of terminal equipment and the analog network, it is important that the modem provide a standard interface to all such terminal equipment. For years the EIA standard RS-232 has provided the specifications for both mechanical and electrical interfaces to the modem. Performance characteristics for this standard and a recently adopted interface standard are given in Table 9.2.1.1-2. The new standard, RS-449, is designed to operate at data rates up to 100 kbits/sec at a distance of 1,200 m in the balanced line mode. For shorter distances of down to 12 m, the interface will accommodate up to 10 Mbits/sec. In the unbalanced mode, the interface allows 3 kbits/sec over a 1,200-m link or up to 300 kbits/sec at a

TABLE 9.2.1.1-1. CHARACTERISTICS FOR TYPICAL STATE-OF-THE-ART MODEMS

MANUFACTURER	MODEL	SYNCHRO-NIZATION	LINE TYPE	OUTPUT DATA RATE (kbits/sec)	BER	COST $
Codex	1960	Synch	Point-to-Point	4.8, 7.2, 9.6	10^{-6}	8,500
Codex	8310	Synch	Point-to-Point	48, 50, 56, 64	10^{-8}	-
Datec	30	Asynch	Switched and Point-to-Point	0.3	-	310
Develcon	DS513	Asynch	Hardwire-Limited Distance	9.6	-	275
General Datacom	201C	Asynch	Switched and Point-to-Point	2.0, 2.4	-	-
General Datacom	2400EP	Synch	Point-to-Point	2.4	-	-
Nesco	420	Asynch	Switched and Point-to-Point	2.4, 4.8, 9.6	-	-
Tele-Dynamics	7208	Synch	Switched and Point-to-Point	4.8	-	-
Vadic	VA3467	Asynch	Switched and Point-to-Point	1.2 each from three modems	-	850
Vadic	VA3415	Asynch	Switched and Point-to-Point	1.2	-	740
Western Union Information Service	4196	Synch	Switched and Point-to-Point	4.8, 7.2	-	6,500

TABLE 9.2.1.1-2. PERFORMANCE FOR RECOGNIZED MODEM INTERFACE STANDARDS

EIA STANDARD	LINE CHARACTERISTIC	SHORT RUN		LONG RUN		CONNECTOR PINS
		DISTANCE (m)	DATA RATE	DISTANCE (m)	DATA RATE	
RS-232C	Unbalanced	15	20 kbits/ sec	-	-	25
RS-449	Unbalanced	12	300 kbits/ sec	1,200	3 kbits/ sec	37
	Balanced	12	10 Mbits/ sec	1,200	100 kbits/ sec	

distance of 12 m. By comparison, RS-232C provided for a 20-kbit/sec link over a distance of 15 m. The provisions of RS-449 allow for an unmodified interface with RS-232 equipment; however, system data rates will be limited to those set by the RS-232 equipment. In addition to the higher achievable data rates, the new interface accommodates seven control and testing functions not previously available. The new specification uses a 37-pin connector of the same type as the 25-pin connector used for RS-232C.

In addition to the mechanical and electrical interface requirements, there must be a compatibility on the language level. Several protocols exist for this purpose, and what is standard depends on the machine characteristics and the location. Byte-oriented protocols are used today by several companies, including IBM, Burroughs, and DEC. These protocols require complex software to separate the message from the overhead. Additionally, they are limited to half-duplex operation. Bit-oriented protocols have a much simpler overhead and require only two characters to identify the beginning and ending of a frame. Several of these bit-oriented protocols exist, including IBM's Synchronous Link Data Control (SLDC), the Advanced Data Communications Control Procedure (ADCCP) of the American National Standards Institute, and the High-Level Data Link Control (HDLC) of the International Standards Organization. The existence of so many "Standards" creates a noncompatibility of equipment. The need for a truly standard protocol exists, but authorities in the area predict that no true standard will come until the turn of the century.

9.2.1.2 Trends and Projected Developments in Modem Technology - The trends in modem technology are based on the implementation of LSI circuitry and microprocessor controllers to reduce the power consumption, size, and cost. They will at the same time increase bit rates, reliability, and error performance. The advances in network technology are dependent on advances in modem technology. The rate for point-to-point leased lines will be limited to a maximum of 14.4 kbits/sec, about 1.5 times the present capability, and these rates, if achieved, will be the result of advances in modem technology in lieu of improvements in distribution networks. Some custom-designed, very-high-cost modems will be capable of operation at higher rates during the early 1980's. Harris Radiation Labs currently has a contract with Rome Air Development Center to develop a modem capable of sending 16 kbits/sec over a single telephone line. The details of specifications and system availability have not been released at this time. Experimental systems operating at 16 kbits/sec have been demonstrated with fair success for digitized speech, but the error rates of these systems are too high for use in data communication. The proper error coding and equalization techniques will make data communication possible for this high throughput rate, but cost of the necessary equipment will be too high in the near future for the average user. The maximum theoretical channel capacity for a voice-grade telephone line is approximately 20 kbits/sec. The present state-of-the-art modems achieve 48% of the theoretical rate, and the 14.4-kbit/sec modems of the early 1980's will achieve 72% of this. The proposed 16-kbit/sec modem will achieve 80% of the maximum and is approaching the practical rate that can be achieved over a voice-grade telephone line. Table 9.2.1.2-1 gives the projected bit rates for modems operating over voice-grade lines in the 1980-1985 timeframe.

The electrical and mechanical interface standards will be consistent with those of RS-449. The agreement on a common protocol standard will be slow, with a true standard unlikely during the 1980-1985 timeframe. Equipment will move toward a standard, but general compatibility will not come until at least the 1990's.

TABLE 9.2.1.2-1. PROJECTED DATA RATES FOR MODEMS OF THE 1980-1985 TIMEFRAME

TYPE OF MODEM	PROJECTED BIT RATE (kbits/sec)	INCREASE OVER PRESENT (%)	PORTION OF THEORETICAL (%)
Multipoint	9.6	100	48
Point-to-Point	14.4	50	72
Special-Purpose	16.0	67	80

Additionally, the networks of the 1980-1985 timeframe will be more distributed than those of today. The average terminal will possess capability to function independently of the network, but reliability of the data link will still be a major concern. The modems of this period will contain automated preventive maintenance functions with automatic switchover to redundancy in the event of a failure. The use of modems will increase as the use of distributed data processing increases. The number of modems in use will be impacted somewhat by the continued expansion of digital network services; however, the expansion of data processing needs is expected to increase at a rate much higher than can be met by digital data services alone.

9.2.2 Front-End Processors

The front-end processor, as discussed in this subsection, is located at the host computer. Its functions are the monitor and control of the communications network-to-host computer interface. In performing this function, there are several tasks performed by the front-end processor. Those classified as basic functions include the control of reception and transmission, responsibility for code conversions, the control of network errors, editing of messages, and the routing of messages to and from the main program processor. In the area of polling and dialing functions, the front-end processor may perform such tasks as the dialing of terminals, answering the line when the computer is dialed, and polling the terminals on either an automatic or programmed basis.

An important purpose of the front-end processor, although not generally counted as a function, is to conserve the amount of processing capability taken from the host computer. To optimize the savings of host computer resources, several other tasks can be assigned on a user-sensitive basis. For example, the front-end processor can control the allocation of a dynamic buffer to minimize the storage required in the main computer. The front-end processor can also serve as a queueing mechanism for both incoming and outgoing messages. Additionally, it can handle scheduling and can allow the CPU to perform other functions during periods when the communications channels are busy. Two other functions that are not always implemented but that should be considered are message-logging and statistics-gathering.

9.2.2.1 State of the Art in Front-End Processors - The front-end processors in general use today are either of a hardwired or programmable nature. The hardwired processors can accommodate a variety of either asynchronous or synchronous terminals. The asynchronous terminals most often accommodated operate at one of seven common speeds: 75, 110, 134.5, 150, 300, 600, or 1,200 bits/sec. The synchronous speeds accommodated include, but are not limited to, 2, 2.4, 3.6, 4.8, 7.2, 9.6, and 19.2 kbits/sec. Higher rates from 48 to 56 kbits/sec are also available with some processors. While the hardwired front-end processors have the advantage of speed, they possess the disadvantage of general inflexibility in the changing of network configuration. Also, the host computer must provide the processing capability for many of the network functions. Despite these constraints, the hardwired front-end processor continues to be used in many systems.

The programmable front-end processor maintains the advantages of both a system interface flexibility and a reduction of the burden on the CPU. Programmable front ends usually incorporate a minicomputer with memory ranging from 4 to 512 kbytes. The functions available on many of the typical units include variable line count, variable speed of the throughput, and adaptive data formatting. These units can also perform error correction, editing, and dialing functions. Options available include statistical multiplexing, data compression, network performance monitoring, and network statistics gathering. The available options vary among manufacturers, and plug-in modules provide a flexibility even above that obtained by programming. With a number of diversified functions available from different manufacturers, a difficulty exists in separating the network control functions from those required for an efficient network interface. Some current systems employ both a front-end processor and a network controller, but many use the front-end processor for both functions. A representative listing of currently used front-end processors and their manufacturers is presented in Table 9.2.2.1-1. The table covers the characteristics of several units currently employed for this purpose.

TABLE 9.2.2.1-1. CHARACTERISTICS OF STATE-OF-THE-ART FRONT-END PROCESSORS

MANUFACTURER AND MODEL	LINES		NUMBER OF HOSTS	WORD SIZE (bits)	COST ($K)	STORAGE MEDIUM	MEMORY SIZE (kbits)	OTHER FEATURES
	NUMBER	RATE (kbps)						
Computer Communications, Inc.								
CC-8	64	230	2	16	88.9	MOS	16 to 128	Peripherals: card reader, line printer, fixed-head disc, movable-head disc
CC-80	960	230	7	16	117.8	MOS	16 to 512	
Counter 3670	384	230	4	16	70 + options	Core	16 to 512	Dynamic line/subchannel assignment, automatic baud rate detection; full-duplex BSC can be used as concentrator or concentrator/controller.
Honeywell 6670	N/A	50	6	18	190	Semiconductor	N/A	Any combination of duplex or full duplex operation
IBM 3705-II	352	56	2	16/32	N/A	Monolithic FET	32 to 256	Can be used as line concentrator
Memorex 1380	240	230	4	16	154	MOS	32 to 512	Direct replacement for IBM 3704/3705. Block-level synchronous scanners and block-level channel adaptors are standard.
Modular Computer Systems, Inc. IV/OP	256	250	1	16	29.5	-	512	Accommodates any type of peripheral
Paradyne Corporation PIX II	N/A	56	N/A	16	15.8	MOS	32	Features include voice adaptor and line switch

The functions available with these systems vary, and the amount of network control depends both on the front-end processor chosen and the configuration of the network. The maximum control is achieved for networks configured with similar equipment, while the least amount of control and flexibility is obtained for terminals with varying characteristics.

9.2.2.2 Trends and Projected Developments in Front-End Processors - The general trend for front-end processing equipment will be for an increase in both flexibility and capability. The major role of the front-end processor is an interface between the central computer and the communications network. The amount of influence that the processor exhibits beyond the interface depends on the user's intentions, but that amount of influence is growing. With the increase in acceptability of the minicomputer, the front-end processor functions have increased to take over those items normally associated with other hardware. In many cases, front-end processors have accepted the role of concentrator and multiplexer on the network side, while relieving the host of the traditional network processing on the other side of the interface. In many cases, the front-end processor will continue to assume more functions, with processors of the future being capable of automatic network restoral in the event of an equipment failure along the line.

The implementation of the distributed network will have an impact on the role of the processor. With passive terminals, the entire burden of the end-to-end communications system is placed with the host and/or front-end processor. With the advent of the intelligent terminal and distributed processing, this changes somewhat. The terminals become more autonomous and the flow of data is filled with more information and fewer control functions. As the functions of formatting and terminal control are further removed from the processor, the possibility for network control is increased. In essence, the front-end processor will begin to look more like a network controller.

The use of hardwired front-end processors will decrease, with programmable processors accounting for more than half of the systems of the early 1980's. The programmable processors will serve to release the host from the burden of control as well as taking control of network functions. The trends and developments in front-end processor technology will generally follow those for processors and memory and are discussed

in Section 7 of this report. As discussed in this section, large computers will exist on a single integrated circuit chip. Thus the front-end processor in 1985 may consist of a single chip built into a box that houses the modem (also a chip and/or a supercomponent), multiplexer, and other special circuitry for interfacing a variety of components (e.g., computers and terminals, etc.) to the communication network.

9.2.3 Data Multiplexers

Data multiplexers, as discussed in this subsection, allow several data terminals to share a communications line. The multiplexer serves to interleave the data of the various remote users so that a more complete utilization can be made of available bandwidth. This can result in reduced line costs, especially when several pieces of low-to-medium data rate equipment are located in a central area and have connections to the same processing unit.

While conventional multiplexers have a definite role in the area of data communications, the newer types of multiplexers with front panel programming and automatic sensing functions will become more important in future data systems. The multiplexer will assume the responsibility for more of the network control functions that were once considered the job of the network controller. These intelligent multiplexers will be a vital part of the distributed network of the future.

9.2.3.1 State of the Art in Data Multiplexers - Line data multiplexers generally use either Time Division Multiplexing (TDM) or Frequency Division Multiplexing (FDM) to interleave the data from several users onto one communications line. The more common technique used today is a TDM system that interleaves the data on a per-bit basis. The multiplexers may operate in either a synchronous or asynchronous mode, with some models offering the capability to mix modes. Typical asynchronous channel input speeds range from 75 to 1,200 bits/sec. Synchronous speeds on a number of units start at 1.2 kbits/sec and include 2, 2.4, 3.6, 4.8, and 7.2 kbits/sec, with some higher-speed multiplexers also accepting inputs at 9.6 kbits/sec. Standard output rates are synchronous and include the rates given above for synchronous inputs. Some higher-rate multiplexers have output rates of up to 64 kbits/sec. Table 9.2.3.1-1 presents the characteristics for several typical units.

Multiplexers in the strict sense (TDM for example) allot a given amount of time (or frequency for FDM) for transmission of data to and from a particular terminal. The result is that a portion of the line is dedicated to the user whether any data are present or not. Therefore, the fixed multiplexing schemes are better adapted for use with systems that have a relatively constant data rate. Systems with "bursty" data are better suited for implementation with a packet switching arrangement, although this scheme may not be suitable for many user applications. An alternative would be to use an intelligent multiplexer, also known as a concentrator. The concentrator does the work of a multiplexer, but it contains buffer storage to prevent data overflow during burst periods. The system then spreads the data out on the line, taking advantage of break periods in the other users' data. Additionally, most concentrators multiplex the data on a block basis rather than on a bit or character basis. A bit code is sent to signify the status of any idle terminal, thus releasing that terminal's share of the line for use by other terminals.

TABLE 9.2.3.1-1. CHARACTERISTICS FOR TYPICAL STATE-OF-THE-ART DATA MULTIPLEXERS

MANUFACTURER	MODEL	NUMBER OF INPUT LINES	INPUT RATES		OUTPUT RATES (kbits/sec)	USER TRANSPARENCY	COST ($)
			ASYNCHRONOUS (bits/sec)	SYNCHRONOUS (kbits/sec)			
IBM	2712-1	10	134.5	None	--	Yes	--
IBM	2712-2	14	74.2	None	--	Yes	--
Codex	920	Up to 64	75, 110, 134.5, 150, 300, 600, 1,200	1.2, 2.0, 2.4, 3.6, 4.8, 7.2	9.6	Yes	1,800 (16 ch) 2,700 (64 ch)
Codex	8014	20	None	2.4 to 38.4 in 2.4-kbit/sec increments	64	Yes	--

As the capability of the equipment at the user's site increases, the system takes on the characteristics of a distributed network. The intelligent multiplexer takes on other tasks besides pure multiplexing of data. The intelligent multiplexer can be used as a network node, connecting to remote job entry equipment in addition to the conventional user terminal. As the intelligent multiplexer gains capability, it can be used for data compression. The simplest data compression techniques of this type reduce the data required to transmit repeated characters. More sophisticated data compression uses codes to assign bit patterns to the various characters. By assigning the simplest patterns to the more frequently used characters, the transmission efficiency can be improved by approximately 50% over that of a standard 8-bit format. Intelligent multiplexers are also capable of providing error correction through the use of an automatic repeat request feature.

The advantages offered by multiplexers vary according to the users' needs. Where many low-rate terminals are in close proximity and communicate with the same processing unit over a relatively long distance, the multiplexer can be justified by reducing the need for having several under-utilized lines. On the other hand, the cost of the multiplexer could not be justified for a small number of terminals (two or three) operating over local lines. Likewise, the advantages are great for the low-rate user, since many low-rate users can be placed on a voice-grade line. This advantage decreases as the user's data rate approaches the capacity of the line.

9.2.3.2 Trends and Projected Developments in Data Multiplexer Technology - The trend in data multiplexer technology is toward an increase in throughput data rates and an increase in control functions. Multiplexers of the 1980-1985 timeframe will have a large number of input ports and these ports will be capable of supporting higher data rates. The output rates of these multiplexers will also increase, with the higher data rates being accommodated by the digital data networks rather than by modems and analog networks. High-rate data multiplexers (handling hundreds of lines and having output rates of up to a few megabits per second) will tie to the small terminal satellite Earth stations proposed for the early 1980's.

At the slower multiplexer rates (9.6 kbits/sec and below), the multiplexers will exhibit increased capability due to internal processing. The smart multiplexer will be common, having the standard functions of data compression and error detection, in addition to offering an automatic adaption to terminal speed and format. These adaptive systems will develop with the increased use of LSI and microprocessor circuitry and will be found in some form in all multiplexer applications except the simplest.

The use of the intelligent multiplexer will become even more important with an increase in the number of all digital networks. Using the system currently being implemented for analog-to-digital conversion, the voice-grade analog line (capable of 9.6 kbits/sec at most) of the telephone company is equivalent to a 56-kbit/sec digital line. To use this capability efficiently, digital lines serving customers would of necessity need a multiplexer to use the line's capacity while operating terminals. Additionally, the standard multiplexer is only efficient if all terminals are operating at rated capacity. The intelligent multiplexer will reduce a great portion of this inefficiency, allowing active terminals to take advantage of the unused portions of the data lines.

REFERENCES - SECTION 9

9-1 H. I. Maunsell, R. B. Robrock, and C. A. von Roesgen, "The M13 and M34 Digital Multiplexers", IEEE 1975 International Conference on Communications, pp. 48.5-48.9

9-2 F. M. E. Sladen and S. R. Norman, "Data Transmission Using Optical Fibers", Electronic Engineering, pp. 57-63, June 1977

9-3 Wilburn L. Pritchard, "Satellite Communication - An Overview of the Problems and Programs", Proceedings of the IEEE, Vol. 65, No. 3, pp. 294-307, March 1977

9-4 "European Satcoms Near Operational Goal", Aviation Week and Space Technology, pp. 113-115, October 17, 1977

9-5 Douglas D. Crombie, "Lowering Barriers to Telecommunications Growth", Data Communications, p. 80, June 1977

BIBLIOGRAPHY - SECTION 9

Bjorn Abrahamsen, Mils Holte, and Terje Roste, "A 560 Mbit/sec Digital Transmission System Over Coaxial Cable", IEEE, Zurich, 1976, pp. A4.1-A4.6

Ira Jacobs and Stewart E. Miller, "Optical Transmission of Voice and Data", IEEE Spectrum, pp. 33-41, February 1977

Leonard Kleinrock, "On Communications and Networks", IEEE Transactions on Computers, Vol. C-25, No. 12, pp. 1326-1334, December 1976

H. K. Pfyffer, "Digital Multiplexing and Line Transmission -- A Survey", IEEE, Zurich, 1976, pp. A1.1-A1.6

P. E. Rubin, "The T4 Digital Transmission System -- Overview", IEEE 1975 International Conference on Communications, pp. 48.1-48.4

F. O. Waldhauer, "A 2-Level, 274 Mb/s Regenerative Repeater for T4M", IEEE 1975 International Conference on Communications, pp. 48.13-48.17

Howard Falk, "'Chipping in' to Digital Telephones", IEEE Spectrum, pp. 42-46, February 1977

James Martin, Introduction to Teleprocessing, Prentice-Hall, Inc., Englewood Cliffs, New Jersey

R. G. Gould and Y. F. Lum, "Communication Satellite Systems: An Overview of the Technology", IEEE Press, 1976

G. David Forney and Robert W. Stearns, "Statistical Multiplexing Improves Link Utilization", Data Communications, pp. 37-41, July/August 1976

Elton Shearman, "Implementing Distributed Networks with SDLC", Digital Design, pp. 60-66, March 1977

Dan M. Bowers, "Data Communications - A System Mentality is Needed", Modern Data, pp. 51-58, April 1975

"Focus on Data Communications", Signetic Microcomputer News, Vol. 1, No. 3, March/April 1977

S. Browne, "Inside the 13-Channel SSB Station Carrier", Telecommunications, pp. 55-60, April 1977

K. C. Kao and M. E. Collier, "Fibre-Optic Systems in Future Telecommunication Network", Telecommunications, pp. 25-32, April 1977

"SBS Clarifies Initial Satellite Filing", Data Communications, pp. 11-12, July/August 1976

"Microprocessors Take Over Execution and Overhead of Full- and Half-Duplex Protocols", Data Communications, pp. 55-56, July/August 1976

Ralph E. DeMent, "Pricing Switched Services", Data Communications, pp. 57-62, May/June 1976

Richard B. Hovey, "Packet-Switched Networks Agree on Standard Interface", Data Communications, pp. 25-39, May/June 1976

Dixon R. Doll, "Relating Networks to Three Kinds of Distributed Function", Data Communications, pp. 37-42, March 1977

Timothy Johnson, "Projecting the Future Roles of Packet-Switching Networks", Data Communications, pp. 18-23, July/August 1976

Miki Takeuchi and Martin Weisbart, "A Worldwide Network Is Formed by Three Independent Centers", Data Communications, pp. 67-74, March 1977

J. Christopher Burns, "The Evolution of Office Information Systems", Datamation, pp. 60-64, April 1977

Gilbert Held, "Sharing the Line: A Cheaper Way than Multiplexing", Data Communications, pp. 79-87, March 1977

James Cody, "Bank Building Net to Serve Statewide Community", Computerworld, p. 39, May 2, 1977

"'Genie' Serving Eastern Airlines Information Needs", Computer World, p. 48, May 2, 1977

Ronald A. Frank, "FCC Tariffs Tymnet as Second Public Packet Carrier", Computerworld, p. 5, April 11, 1977

George R. Davis, "Earth Stations Slim Down to Meet Grass-Roots Demands", Microwaves, pp. 36-46, January 1977

Howard L. Crispin, "Earth Stations Dot the Globe", Microwave Systems News, pp. 30-33, March 1977

Louis Cuccia, Wasson Quan, and Carl Hellman, "Above 10 GHz Satcom Bands Spur New Earth Terminal Development", Microwave Systems News, pp. 37-57, March 1977

Stephen A. Kallis, "Networks and Distributed Processing", Mini-Micro Systems, pp. 31-40, March 1977

C. E. White, "Packet Switching for Fast and Secure Data", Telecommunications, pp. 33-39, January 1978

J. L. Holsinger, "Where Are Modems Going?", Telecommunications, pp. 12-15, June 1977

E. R. Kretzmer, "The Evolution of Techniques for Data Communication Over Voiceband Channels", IEEE Communications Society Magazine, pp. 10-14, January 1978

Glenn Hartwig, "Network Control: Managing the Data Environment", Data Communications, pp. 41-52, December 1977

Dick Hackmeister, "Bit-Oriented Protocols Plus Standard Interfaces Ease Data Communication", Electronic Design 12, pp. 36-40, June 7, 1977

Ray Sarch, "What Intelligent Time-Division Multiplexers Offer", Data Communications, pp. 29-36, July 1977

"RS-449 Ousts RS-232", Digital Design, p. 18, December 1977

"A Look at Message/Packet Switching", Telecommunications, pp. 44-52, January 1978

Martel Firing, "Modems: A Maturing Market", Telecommunications, pp. 47-50, June 1977

"Businesses to Test Satellite Communications", Electronic Design 25, pp. 21-22, December 6, 1977

George S. Li, "Analysis of Submarine Cable and Communication Satellite Systems Reliabilities", Data Communications, p. 84, August 1977

"Innovative Limited-Distance Mode", Telecommunications, pp. 58-59, June 1977

Richard Stuart, "Advancement in Data Set Design", Telecommunications, pp. 42-44, June 1977

H. C. Folts and I. W. Cotton, "Interfaces: New Standards Catch Up With Technology", Data Communications, pp. 31-40, June 1977

"New Products", Data Communications, p. 60, July 1977

J. P. Martino and R. C. Lenz, Jr., "Broadcast Satellite Communications Forecast", Telecommunications, pp. 33-36, October 1977

"Western Union Plans a DDS, Too", Data Communications, pp. 22-23, June 1977

"SBS Chooses Hughes Satellite", Electronics, pp. 32-33, October 13, 1977

"SBS' Controller Costs Are Going Into Orbit", Data Communications, p. 9, July 1977

"Muzak to Send Music by Satellite?", Electronics, p. 33, January 20, 1977

"Bell's Chicago System Goes Like Gang Busters", Electronics, pp. 43-45, December 22, 1977

"Telephone Transmission Via Optical Fibers Begins Regular Service", Iron Age, p. 40, June 27, 1977

"DSDS Ready in 4th Quarter", Data Communications, pp. 14-15, July 1977

"CODEC Meets Bell μ-255 Specs", Electronic Products, p. 57, January 1978

H. L. Van Trees, E. V. Hoversten, and T. P. McGarty, "Communications Satellites: Looking to the 1980s", IEEE Spectrum, pp. 43-51, December 1977

"Arpanet Concept Used at Citibank for Dedicated Net", Data Communications, pp. 13-18, June 1977

"Small Antenna Ruling is Causing Mini-Market Boom", Microwave Systems News, pp. 75-80, March 1977

"Small Stations OK'd for Satellite CATV", Electronics, pp. 33-34, January 6, 1977

R. A. Frank, "Document Reveals Bell Plans Packet Network", Computerworld, pp. 1-2, September 5, 1977

"The Trend: Digital Switching", Telephone Engineer and Management, pp. 48-55, November 15, 1977

John S. Mayo, "The Role of Microelectronics in Communication", Scientific American, pp. 192-209, September 1977

"Bell Eyes Digital Spot-Beam System", Electronics, pp. 32-34, December 22, 1977

"British Hope to be First with Telephone Fiber Optics", Microwaves, p. 17, June 1977

"Dual Mission Complicates Data Relay Satellite Work", Aviation Week and Space Technology, pp. 155-157, October 17, 1977

L. P. Yeh, "Satellite Communications and Terrestrial Networks", Telecommunications, pp. 23-30, October 1977

Michel C. Jeruchim, "A Survey of Interference Problems and Applications to Geostationary Satellite Networks", Proceedings of the IEEE, Vol. 65, No. 3, pp. 317-331, March 1977

"Modem Characteristics Survey", Telecommunications, pp. 53-56, June 1977

P. D. Shaft and J. A. Roberts, "Optimum Allocation of Multibeam Communications Satellite Resources", IEEE Transactions on Communications, pp. 1195-1200, October 1976

C. Louis Cuccia, "Japan Bids for Leadership in Space Communications", Microwave Systems News, pp. 57-73, March 1978

Leif Lundquist, "TDMA: The Answer to European SATCOM?", Microwave Systems News, pp. 79-86, March 1978

10. INFORMATION PRESENTATION ELEMENTS

Information presentation elements are discussed in this section in terms of the dynamic or real-time presentation elements and the hardcopy presentation elements that are depicted in Figure 10-1. Dynamic presentation elements are those that are capable of being driven and/or changed in real time, in contrast to the hardcopy elements, which are permanent in nature once they are produced. For purposes of classification, dynamic presentation elements include certain classes of projection systems that use hardcopy media (e.g., film) as the image source. Thus, for this example, the systems that generate the hardcopy image source are classified as hardcopy presentation elements, and systems that project the image are classified as dynamic presentation elements.

For more than a decade, the information presentation/display market has been projected as ready to explode. Indeed, the market has "exploded" in the area of alphanumeric terminal devices, but has been less successful in the area of graphics than forecasted. To a certain extent, advancements in technology have been reflected by this market. Many of the most significant advances that have occurred in displays recently have been the result of advances in supporting technologies, primarily integrated circuit technology. Among the significant advances that have occurred as a result of IC technology are the intelligent terminals, graphic displays that are refreshed from solid-state memories, and digital control of graphic functions.

Display devices still employ the cathode ray tube (CRT) as the dominant display medium, and indications are that this will not change significantly during the next several years; however, other technologies are beginning to emerge as technically feasible approaches. The application of these technologies will probably be limited for the foreseeable future to special requirements that can justify the additional cost over that for a CRT-based system. At present, no technology is cost-competitive with the CRT, although certain classes of plasma panels are showing progress in that direction.

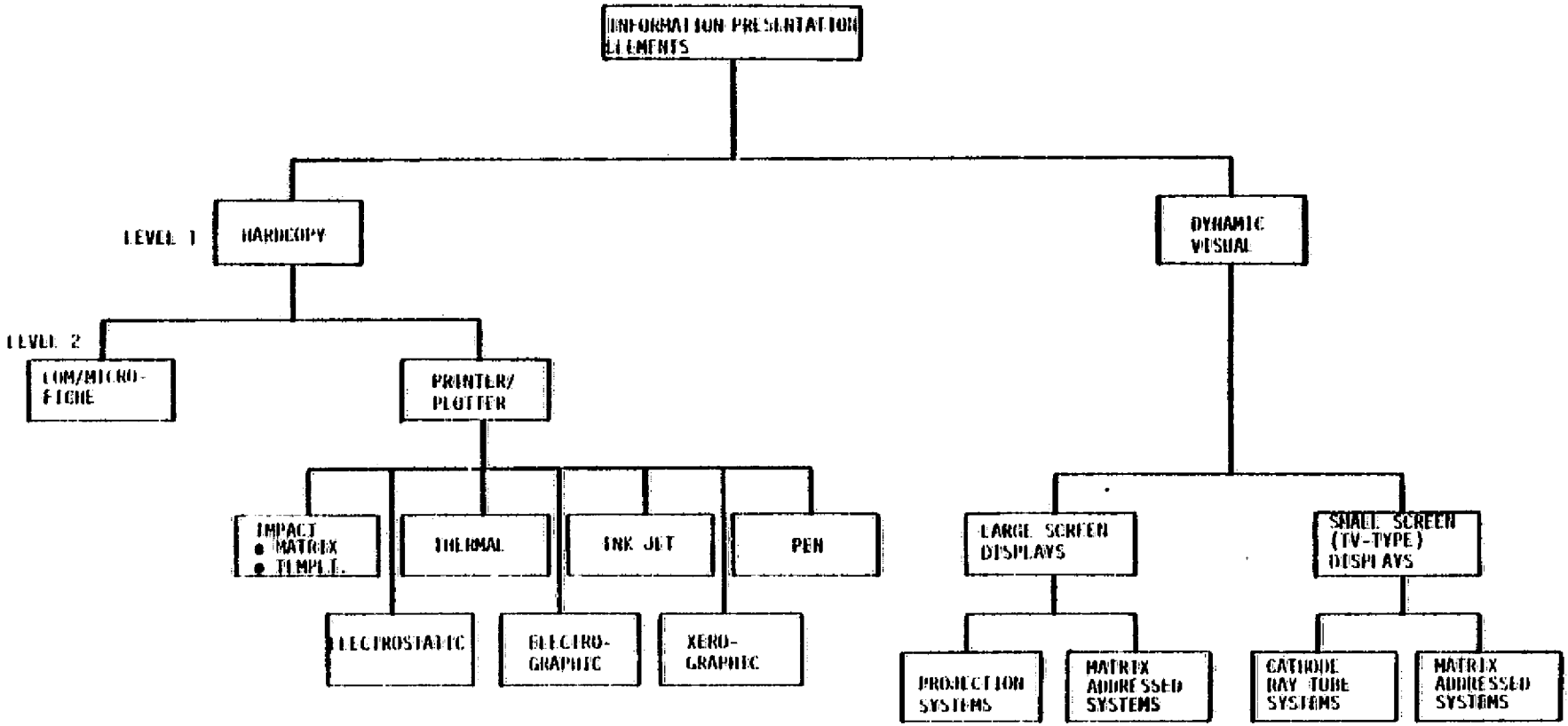


FIGURE 10-1. INFORMATION PRESENTATION ELEMENTS

Research is going on in many areas of display technology at the present time. In fact, so many areas are being investigated that the efforts appear to be fragmented. This fragmentation clearly disperses the available funding to the point that it probably restricts technology development. The one factor that places the most pressure on manufacturers to improve display media technology is the search for higher-performance home television/home entertainment displays to replace the current CRT systems. A breakthrough in one or more areas may result in more funding being diverted to that area, which could conceivably produce a system that is cost- and performance-competitive with the CRT. However, most authorities are not presently willing to commit themselves as to which display technology will experience this first breakthrough, or indeed if such a breakthrough occurs, whether that technology will win the final competition as the predominant display medium for computer-generated data, home television, and other information presentation functions.

10.1 DYNAMIC INFORMATION PRESENTATION ELEMENTS

Dynamic information presentation elements include the visual presentation systems and represent the most important method for presenting dynamic data. Visual systems are classified as either small-screen (TV-type) or large-screen systems. Small-screen systems are generally classified as those devices that have display screens that measure 25 in. or less diagonally. This distinction has been primarily the result of practical limitations on the size of cathode ray tubes, which are the principal dynamic small-screen display medium. This distinction between large- and small-screen displays may change as technology advances.

Technology is discussed in the following subsections in terms of two specific aspects of display technology: 1) features that are provided independent of the display medium (e.g., intelligent terminal features) and 2) the display medium. Many of the features available on terminal devices are not limited by technology but are functions of human factors and cost.

10.1.1 State of the Art in Small-Screen Displays

The state of the art in small-screen (TV-type) dynamic information presentation elements is discussed in terms of features available on alphanumeric terminals, graphic display terminals, and image presentation devices. The specific display device technologies (e.g., CRTs, plasma panels, LCDs) are discussed and compared in subsequent paragraphs.

Alphanumeric display terminals perform as remote communication terminals (similar to the hardcopy teletype devices) and as local terminals for interactive operations (e.g., editing and data base access) with computers. As such, these devices are available as 1) intelligent terminals (possess memory and limited programming and processing capability), 2) smart terminals (possess editing capability), and 3) dumb terminals (strictly dependent on the host processor for all intelligence). The dumb terminal has been in use since the mid-sixties, and although it still serves the needs of numerous users, many of the newer terminals are incorporating microprocessor-based intelligence. Some industry officials feel that the emergence of low-cost microprocessor-controlled terminals in 1977 was most significant (Ref. 10-1) in terms of reducing the cost of display devices. The introduction of microprocessors into terminals allowed some manufacturers to lower their parts count by 25 to 30% for dumb terminals and by 30 to 35% for smart terminals. The resulting decrease in the price of display terminals will serve to reinforce the increasing sales of microprocessor-based terminals. The majority of display terminals in service do not have microprocessor controllers, but an increasing number of new terminals are equipped with microprocessors.

A characterization of the state of the art for alphanumeric displays, excluding the display medium itself, is difficult to accomplish because certain characteristics are more a function of human factors and cost than of technology. The primary technology items are features of intelligent terminals, such as edit capabilities, memory size, and stand-alone processing capabilities. Tables 10.1.1-1 and 10.1.1-2 present the characteristics and features, respectively, for

TABLE 10.1.1-1. CHARACTERISTICS OF TYPICAL ALPHANUMERIC DISPLAY TERMINALS

CHARACTERISTIC	RANGE
Display Screen Viewing Area	50 to 80 in^2
Number of Display Lines	24 typical; up to 40
Number of Characters/Line	80
Maximum Viewable Characters	1,920 typical; up to 3,200
Character Set	64, 96, 128
Character Code	ASCII
Symbol Formation	5 × 7 dot matrix
Keyboard	Typewriter
Transmission Mode	Half/full duplex
Transmission Rate	2,400 to 9,600 bits/sec
Transmission Technique	Synchronous or Asynchronous
Communication Protocol	ASCII/BSC/SDLC
Message Format	Character or Block
Interface	EIA RS-232C
Display Monitor	CRT

TABLE 10.1.1-2. TYPICAL ALPHANUMERIC DISPLAY TERMINAL FEATURES

FEATURE	COMMENTS
Memory	Available to 64K words
Microprocessor Control	Widely available
Edit Features	
Cursor Positioning	U, D, L, R, H on most
Cursor Blinking	Available on most
Character Insert/Delete	Available on most
Line Insert/Delete	Widely available
Erase	Character, line, screen common
Roll	Limited availability
Paging	Limited availability
Color	Limited availability
Auxiliary Storage	Limited availability
Light Pen/Other Select Modes	Limited availability
Selectable Brightness	Up to 2 levels fairly common
Compatible With One or More Standard Devices	Generally available
Internal Refresh	Trends in direction of Internal Refresh

typical alphanumeric terminals available on today's market. These capabilities and features are not considered to be the limit of technology, but are representative of the features and capabilities that are available at a cost that the market will bear.

The CRT still remains the principal display medium for alphanumeric display terminals. Plasma panels are finding limited use for special applications, but they are not cost-competitive with the CRT. Other technologies (e.g., LEDS, and liquid crystals) are receiving limited use in applications that are generally limited in the number of characters and that are no more than one line. Several companies are developing small-screen TV-type displays using liquid crystals, LEDs, and electroluminescent displays, but these systems have limited applications in their current configurations and lack the resolution necessary for most terminal alphanumeric and graphic display applications.

The CRT is also the primary display medium used for graphic display applications. For these applications, display size is more important than it is for alphanumeric display applications, and systems measuring up to 25 in. diagonally are available. Graphic displays are available in both black and white, with from 2 up to 64 shades of gray, and color. The resolution for both types typically goes up to 1,024 lines for raster scan systems and up to 4,096 for stroke writing display systems. At least three companies (Adage, Vector General, and Evans Sutherland) build graphic display systems that produce dynamic perspective drawings of complete three-dimensional objects. These systems are capable of maneuvering and/or manipulating the displayed picture through offsets, rotation, translation, scaling up, scaling down, windowing, clipping, and zooming. Most of the new systems perform these functions digitally. Prior to the past 2 or 3 years, these functions were performed using analog techniques.

Table 10.1.1-3 presents state-of-the-art characteristics and features for graphics display systems. This table presents typical features that are available in various systems, although no individual system provides all of the features listed. Table 10.1.1-4 gives a representative

TABLE 10.1.1-3. STATE-OF-THE-ART GRAPHIC DISPLAY SYSTEM CHARACTERISTICS/FEATURES

CHARACTERISTICS	COMMENTS
Display Media	CRT
Maximum Viewable Area	208 in^2
Addressable Matrix	4,096 × 4,096 typical; 8,192 × 8,192 maximum
Viewable Matrix	1,024 × 1,024 typical; 4,096 × 4,096 maximum
Word Length	16 bits
Points/sec	10^6
Vectors/sec	10^6 in/sec
Characters/sec	330,000
Character set	128 character ASCII set typical
FEATURES	**COMMENTS**
Color	Available on some units
Intensity/Shades of Grey	Available on some units
Perspective (3-D) Views	Available from 3 vendors
Translation/Rotation/Offset	Widely available
Line Texturing	Generally available
Windowing/Zoom	Generally available
Conics	Available on some units
Refresh Mode	Trend to internal refresh
Internal Processor	Trend to internal processor
Source Language	Wide use of FORTRAN

TABLE 10.1.1-4. TYPICAL GRAPHIC DISPLAY SYSTEMS (UP TO 1,024 MATRIX)

MANUFACTURER	MODEL	VIEWING AREA (in.)	ADDRESS MATRIX	VIEWING MATRIX	WRITING SPEED (Vect/sec)	CHARACTER SPEED (Char/sec)	COLOR CAPABILITY
Control Data Corporation	240	21 Diag	1,024 × 1,024	1,024 × 1,024	28K	178.5K	No
Digital Equipment Corporation	GT15	NA	8,192 × 8,192	1,024 × 1,024	33K	69K	No
Evans and Sutherland	Pic. System 2	21 Diag	4,096 × 4,096 × 64	--	50.7K	210K	Yes
Hewlett-Packard	2648A	5 × 10	360 × 720	360 × 720	--	--	No
IBM Corporation	2250	12 × 12	1,024 × 1,024	1,024 × 1,024	10K to 85K	84K	No
Imlac Corporation	PDS-16	12 × 12	1,024 × 1,024	1,024 × 1,024	9K in/sec	1,450	No
Information Displays, Inc.	IDI-10	17 Diag 21 Diag	1,024 × 1,024	1,024 × 1,024	5.9K	83K	No
Megatek Corporation	6000	19 Diag	1,024 × 1,024	1,024 × 1,024	178K	35.6K	No
Ramtek Corporation	FS-2000	STD CRT Monitor	256 × 256	256 × 256	--	31K	Multicolor
Sanders Assoc.	900	24 Diag	4,096 × 4,096	1,024 × 1,024	25K	250K	No

listing of graphic display systems having up to a 1,024 by 1,024 viewing matrix. Several ofrer color monitor capability, and one model in this group offers a three-dimensional address matrix for generating perspective views. Table 10.1.1-5 is a representative listing of manufacturers having models with greater than a 1,024 by 1,024 viewing matrix. Again, several models offer the option for color display, and three models have a three-dimensional address matrix. Several of the models listed have a viewing matrix of up to 4,096 points, but this is the upper limit for a practical system with current technology.

Most of the display monitors in use today use the raster scan technique for illuminating the CRT. This method employs an electron beam to cover the screen with a series of parallel lines. Amplitude modulation of the beam produces the display pattern. The image on the tube needs to be refreshed at least 50 times per second if the display is to be free of flicker. To do this, the image is stored in a digital memory and the information is continuously cycled to the display. An alternate method uses the vector (stroke writing) CRT, in which the electron beam draws each image line on the screen. The vector stroke image must also be refreshed at least 50 times per second. Some of the advantages of the vector systems are that the line drawings are very straightforward, and lines can be added, moved, or deleted easily. There is no distracting raster in the background, so the lines are bright and the contrast is high. Additionally, the resolution is good, with up to 4,096 by 4,096 points available. Color capability is provided in stroke writing vector display systems by using a special type of CRT known as the beam penetration tube. The beam penetration tube differs from the standard CRT in that the color phosphors are deposited on the screen in layers rather than as discrete dots. The layer (or color to be displayed) is selected by the amount of energy (applied acceleration voltage) in the electron beam that is swept across the screen. The current technology beam penetration tube systems do not provide color blending capability (i.e., each vector must be a discrete color) and the number of colors that can be displayed is the same as the number of phosphor layers.

TABLE 10.1.1-5. TYPICAL GRAPHIC DISPLAY SYSTEMS (GREATER THAN 1,024 MATRIX)

MANUFACTURER	MODEL	VIEWING AREA (in.)	ADDRESS MATRIX	VIEWING MATRIX	WRITING SPEED (Vect/sec)	CHARACTER SPEED (Char/sec)	COLOR CAPABILITY
Adage, Inc.	GS/330	15 × 17	8,192 × 8,192 × 8,192	3,000 × 3,400	57K	142K	4 Color (Opt.)
Control Data Corporation	777-2	21 Diag	4,096 × 4,096	4,096 × 4,096	10K	30K	No
Hughes Aircraft Company	Graphic Display Terminal	9 × 12 11 × 15 (Opt.)	8,192 × 8,192	2,048 × 1,536	1,760 in/sec	2K	No
Imlac Corporation	PDS-4	10 × 10 10 × 12 (Opt.)	2,048 × 2,048	2,048 × 2,048	80K in/sec	2.6K	No
Lundy Electronics and Systems, Inc.	32/370	16 × 13	2,048 × 2,048	2,048 × 2,048	1,000K in/sec	167K	No
Lundy Electronics and Systems, Inc.	2600	21 Diag	4,096 × 4,096	4,096 × 4,096	1,000K in/sec	286K	4 Color (Opt.)
Sanders Assoc.	7000	24 Diag	2,048 × 2,048	2,048 × 2,048	25K	250K	No
Tektronix, Inc.	4010	15 × 11	4,096 × 4,096	4,096 × 3,120	5K in/sec	4K	No
Vector General	Series 3	13 × 14	4,096 × 4,096 × 4,096	4,096 × 4,096	39K	133K	No
Vector General	3404	19 Diag	65K × 65K × 65K	4,096 × 4,096 × 4,096	55K	180K	4 Color (Opt.)

One modification of the CRT that has come into widespread use for graphic display systems is the direct view storage tube (DVST). This tube uses a bistable phosphor on the screen. The main electron gun activates the phosphor, much as in a standard CRT. Several auxiliary electron guns then flood the screen with enough energy to keep the phosphor energized. This allows a flicker-free image to be realized without the need for internal refresh memory. Since the DVST is compatible with the vector writing techniques, high resolutions of up to 4,096 by 4,096 can be realized. A DVST can retain an image for up to several hours, but has the disadvantage of storing the entire image. Lines cannot be erased or altered on an individual basis and any change requires that the entire screen must be erased and rewritten.

Another method of obtaining storage is with a scanned storage tube (SST). An electron beam scans a selenium-coated target in a vacuum tube to write or erase lines of the display image. The image is stored on the coated target and is read by scanning the target with a second beam. Since the picture is stored in analog form, features such as zooming and windowing become easily implemented. Good resolution can be obtained from this method, but is dependent on the scanning method used to read the storage target. Many systems use a standard TV raster scan, which provides flexibility but degrades resolution.

A refresh and storage method that is gaining widespread acceptance for graphic display devices is the internal solid-state memory. Most of the systems presently employing this method use MOS technology for the memory devices. Image processing (such as zooming and windowing) can be performed by digital processing. Many systems provide an addressable matrix that is several times larger than the viewing matrix, making these features possible.

Information presentation devices that display images in lieu of alphanumerics or graphics play at least two roles for presenting high resolution images. The traditional role is that of the direct-view monitor, similar to the home television set. This type of monitor is

available commercially in either color or shades of gray with display screen sizes to 25 in. diagonally and 1,024 TV lines resolution. The limitation on this type system is one of bandwidth, which is a function of the scan and refresh rates. Increasing the resolution in each dimension causes the bandwidth to increase as a squared function; thus doubling the resolution will cause the bandwidth to increase by a factor of four. This soon becomes prohibitive unless the refresh rate can be reduced, which causes flicker. Thus, to take full advantage of the potential resolution capabilities of the CRTs that are presently available, the CRT should be used in a manner that eliminates the need for refreshing, and thus direct viewing. The approach is to scan the CRT in a slow mode and let the light created by each scan line expose a film where the image is recorded. One commercially available system produces a 70-mm film from a 4,600-line CRT with a 3- by 3-in. raster. The CRT uses a 0.0006-in. CRT light source. Linearity of the system is 0.05%; short-term stability is 0.001%; and long-term stability is 0.0005%.

10.1.1.1 Cathode Ray Tubes (CRTs) - Since the development of the CRT in the early 1930's, it has been the dominant (and until recently, the only) medium for display applications requiring a large number of picture elements. The CRT has advanced from its original application as an oscilloscope display to the capability for high-resolution color TV systems. During the past 15 years, it has gained extensive popularity for the display of computer-generated images, graphics, and alphanumerics.

CRTs are widely available in color, black and white, and shades of gray. The largest color CRTs produced commercially have display screens that measure approximately 25 in. diagonally, with a viewing area of slightly more than 300 in^2. While this is not an absolute limit in size, any increase beyond this causes rapid increases in volume, weight, and cost. For sizes larger than 25 in., equivalent brightness and resolution are extremely difficult, if not impossible, to achieve. Table 10.1.1.1-1 presents representative characteristics for state-of-the-art color picture tubes. All size-dependent characteristics are normalized to the values for a 25-in.-diagonal tube.

TABLE 10.1.1.1-1. TYPICAL SHADOW MASK COLOR TUBE CHARACTERISTICS (Ref. 10-2)

CHARACTERISTICS	COMMENTS
Maximum Picture Area	315 in^2 (2,032 cm^2)
Peak Brightness	1,000 ft-L
Average Brightness (1.5 mA)	150 ft-L
Luminous Efficiency (Tri-Color White)	
Light Out/High Voltage Power	8 lm/W
Light Out/Total Display Power	3 lm/W
Contrast	50:1
Resolution	
Low Current	400 to 500 TV lines
High Current	100 to 200 TV lines
Minimum Number of Electrical Inputs	8
Total Power Input	100 W
Display Volume (Tube, Yoke, etc.)	0.8 ft^3 (23 liters)
Display Weight	50 lb (23 kg)
Display Cost (OEM)	$90

CRT development has advanced to the point where most viewers are unable to distinguish any improvement in performance other than size. The one characteristic other than size that is distinguishable is resolution, particularly in picture highlight areas. The highest resolution of the raster-scanned systems that are commerically available is 1,024 lines (at 30 frames/sec refresh), and many of these systems fail to meet this specification on the fringes and in the picture corners. A number of vendors are currently working to improve resolution, and improvements should be forthcoming in the near future. As was pointed out, vector writing systems are capable of much higher resolution (up to 4,096 lines). Although this resolution is beyond the limits of perception of the eye, it can be utilized in such applications as the film exposure system previously mentioned.

Another area that has been studied for some time concerns the development of a flat CRT. In a conventional CRT, the electron gun must be set in the neck of the tube, behind the screen. The depth of the tube (length of the neck behind the screen) is determined by the amount of deflection that can be obtained. Numerous efforts have been made to reduce or eliminate the neck, and this has been successfully accomplished in a laboratory prototype model. Work at Texas Instruments has resulted in a 6- by 8-in. CRT that is only 2 in. deep. It is a digitally addressed cathodoluminescent display capable of presenting 1920 alphanumeric characters. Each character is formed in a 5 by 9 dot matrix. A multiple-layer grid system is used to cut the number of input leads from 590 (on an equivalent plasma panel display) down to 108. In addition, the flat CRT has a spot brightness of 500 ft-L. This display tube is still in the development stages and it is not known when it may become available on the commercial market.

10.1.1.2 Matrix-Addressed Panels - Almost all flat panel displays (excluding projection systems, which for purposes of this technology assessment are classified as large-screen displays) use a matrix of

display elements. The matrix-addressed devices discussed in this section are classified as small-screen displays since all such systems are currently in the size range of the CRT. However, some of the technologies discussed may achieve the size of large-screen displays in the future. Also, some existing large-screen displays are matrix-addressed panels.

Three of the matrix-addressed techniques that appear to offer high potential for use as TV-type displays are plasma panels, electroluminescent panels, and liquid crystal panels. There are many other technologies under laboratory investigation, and these cannot be ruled out. Although some of these technologies are only suited for static display applications, others may prove suitable for TV-type applications.

10.1.1.2.1 Plasma Panels - Plasma panels (sometimes referred to as gas discharge panels) are the most advanced of the matrix-addressed panels and come closest to competing with the CRT for computer-driven displays. Plasma panels are classified as either dc gas discharge devices or ac gas discharge devices. Both of these techniques are commercially available as flat panel displays and both are receiving attention from a number of companies.

Unlike the CRT, the electrodes of the gas discharge tube are cold and there is no thermionic emission involved. The potential across the electrodes is typically 150 to 170 V. Light coming from the plasma tube is generated by two effects. The first is an ionization referred to as "cathode glow". This cathode glow is the visible output from a number of display devices, including the "nixie" tube and the seven-segment gas display tubes. A larger region of glow space is generated by the "positive column". The positive column is suppressed in many of the devices using cathode glow, but other devices rely on the positive column for their output. The familiar fluorescent lamp uses the positive column glow of mercury to produce ultraviolet light that in turn excites a phosphor coating. The most efficient of the plasma display panels built to date (up to 3.4 lm/W) have incorporated positive column glow, but they must be operated at a high temperature (80 to 85°C), and to date no commercial positive column displays have been marketed.

The dc plasma panels have the electrodes immersed directly in the gas and are driven by unidirectional pulses. The voltages associated with the plasma tubes are too high (170 V) for direct switching with solid-state devices. This can be overcome, however, due to the threshold effect of the plasma tube. The dc voltage is held at a value slightly lower than the firing threshold. A moderate swing of less than 60 V is used to switch the tube into conduction. A pulse of opposite magnitude will cut the tube off. The dc plasma panels usually incorporate a dot matrix format. A 16-character display typically uses 95 columns by 7 rows, for a total of 665 dots. This approach requires a large number of drivers (102) for a conventional matrix driving scheme. An alternative approach using self-scanning techniques reduces the number of display drivers to 10, for a 16-character display. This technique is used in the Burroughs "self-scan" panel. Typical self-scanning panels are available in sizes up to 12 lines of 40 characters. Since the self-scan dc panels do not have an inherent memory, it is necessary to continuously regenerate the display. This poses no problems, except that it limits the maximum number of columns that can be scanned without a noticeable flicker. Currently, the practical size of a scanned dc plasma panel is limited to 200 columns. The dc plasma panels are capable of generating color displays by using ultraviolet light from the glow to excite various phosphors. The choice of colors is limited, however, and there are no color-selective dc displays on the commercial market at present.

The ac plasma panels are energized with bidirectional pulses. The electrodes are insulated from the gas by a thin dielectric film, and the electrode configuration is symmetrical. The dielectric film acts as a capacitor, and charge is stored at each of the electrodes. The initial firing pulse must be large enough to overcome the threshold of the gas cell. The dielectric capacitor charges, and on the next half cycle, adds to the sustaining voltage to continue to light the cell. This process continues until an extinguishing pulse is applied and the capacitance is discharged. Without the capacitor voltage, sustaining pulses will no longer fire the tube. This process gives the ac panel a display memory

and eliminates the need for a dynamic refresh memory. Since the ac plasma panel is a bistable device, it can achieve a high contrast ratio but has no gray scale. Several methods are currently being studied for implementing a gray scale. One method involves the use of several gas elements to make up a picture element. By selectively lighting or extinguishing the dots within each picture element cluster, several shades of gray can be achieved. Another method involves varying the sustaining voltage so that the cell is fired in a degenerative manner and extinguishes itself after a number of cycles. This offers an excellent gray scale but results in low output and poor efficiency.

Much of the current research and development work being conducted for plasma panels is for TV-type image displays. In fact, many of the researchers are television manufacturers. Table 10.1.1.2.1-1 summarizes the current research efforts for gas discharge television displays. The common method of illumination in all cases is to excite a phosphor with the ultraviolet light from the gas discharge. Both Bell Laboratories and Zenith have systems that use a modified Burroughs "self-scan" panel to obtain high-quality pictures. The Bell technique employs the standards of their "Picturephone", while the Zenith system uses commercial television standards. The Hitachi system is similar except it uses a panel that allows greater coupling between the cathode glow and the phosphor dots. The Sony design employs cathode glow to excite a color TV picture, but their scanning is done entirely with digital circuitry. Technical results are promising, but a full color television display requires 750,000 driver transistors.

Zenith, General Telephone and Electronics, and Phillips all have had recent research programs for developing positive column glow plasma display systems for commercial television applications. The positive column techniques are promising and provide greater efficiencies than can be obtained with cathode glow. The panel must be operated at an elevated temperature (80°C), however, to provide sufficient mercury vapor pressure. The positive column techniques are not as refined as those for cathode glow panels, but research is continuing in this area.

TABLE 10.1.1.2.1-1. RECENT RESEARCH IN GAS DISCHARGE TELEVISION DISPLAYS

RESEARCH COMPANY	EXCITATION MODE	SCANNING TECHNIQUE	COMMENTS
Bell Laboratories	Cathode Glow	Self-Scan	Picturephone standards
General Telephone and Electronics	Positive Column	Self-Scan	
Hitachi	Cathode Glow	Self-Scan	Glow well coupled to phosphor dots
Phillips	Positive Column	Self-Scan	
Sony	Cathode Glow	Digital-Scan	Simple display, complex drivers
Toshiba	Constricted Glow	--	High luminance
Zenith	Cathode Glow	Self-Scan	
Zenith	Positive Column	Self-Scan	3.4 lm/W

Toshiba is working on a technique whereby the positive glow discharge is constricted by an insulated aperture. The luminance for this technique is high and the glow can be shifted from electrode to electrode for scanning. Small television pictures are currently practical with this design, but the scan rate is presently below that needed for commercial television.

The future applications of the ac plasma panel may be determined more by supply than by technology. Owens-Illinois, manufacturer of the popular "Digivue" panel, recently announced that they are discontinuing their line of plasma panels. Several companies are currently negotiating for the rights to manufacture the panels, but industry sources are not certain whether the prospective manufacturer will offer the panels on the open market. Fujitsu offers a similar plasma panel that can replace the Digivue in some applications, but users of the panels have not completed full evaluation of the second-source product.

10.1.1.2.2 Electroluminescent Display Panels - Electroluminescent (EL) panels produce light by exciting an electroluminescent phosphor with an electric field. The inherent thinness of these panels offers an excellent potential for the true "flat display". The four major types of EL panels currently in use or under development are: ac powder, dc powder, ac thin-film, and dc thin-film. Table 10.1.1.2.2-1 gives the characteristics for each of these types of panels. For display panel applications, the EL panels generally exhibit a relatively low luminescence, a low luminance efficiency, and a high voltage drive requirement that is difficult to achieve with integrated circuits.

Both the ac and dc panels can be matrixed in arrays with as many as 50,000 elements, but none of the materials has the desired nonlinearity nor is the color gamut good enough for television applications. Additionally, the dc panels have a short life (100 to 300 hr). In terms of life and efficiency, the ac panels are the most favorable. The development of ac thin-film panels is currently receiving much attention in both the United States and Japan. The ac thin-film panel gives excellent contrast

TABLE 10.1.1.2.2-1. STATE-OF-THE-ART CHARACTERISTICS FOR ELECTRO-LUMINESCENT PANELS

TYPE	LUMINANCE EFFICIENCY (lm/W)	LUMINOUS OUTPUT (ft-L)	COLOR	LIFE TO HALF BRIGHTNESS (hr)	COMMENTS
ac Powder	1 to 2	1,000 peak with short life	Green and yellow	5,000	Poor nonlinearity
dc Powder	0.07	10	Yellow	300	Short life
ac Thin-Film	0.1	30 to 60	Yellow-orange	20,000	Receiving much of current research
dc Thin-Film	0.4	35	Yellow	100	Very short life

(yellow-orange against black) and has a long life (20,000 hr). The color of the light can be changed by adding certain rare earth phosphors to the film, but efficiency and life degenerate. Additionally, the ac thin-film panel has been found to exhibit a storage characteristic, a phenomenon not found in other devices. Most of the research on dc panels is occurring in England. The Phosphor Products Company of England has announced a modified technique whereby a dc electroluminescent panel achieves a lifetime of 10,000 hr. The company has built several 1,250-character displays for British military purposes.

Several corporations have research work in progress to build a video display using EL panel technology. One of the larger panels is a 6- by 6-in., 12,000-element electroluminescent panel developed by Westinghouse Research Laboratories. The display uses a thin-film integrated circuit with 24,000 transistors (2 for each element) overlaid with an ac powder electroluminescent phosphor. A similar panel is a 108 by 81 element matrix built by Sharp Corporation. This panel has a picture size of 48 by 36 mm. The Sharp display has an orange-yellow color, maintains eight levels of gray, has a contrast ratio of 50:1, and has a maximum brightness of 60 ft-L.

10.1.1.2.3 Electrophoretic Display Panels - The electrophoretic display, unlike those previously discussed, is a passive display. It functions on the principle of charged pigment particles being transported through an opaque fluid by means of an electric field. The small particles are placed in a colloidal suspension of a contrasting color. When a positive charge is placed on the segment electrode, the display will appear to be the color of the dyed suspending medium. If the segment electrode is made negative, however, the observer will view the charged pigment particles at the front of the display panel.

One of the important features of the electrophoretic display is the ability to store information without the continuous application of an electric field. Electrophoretic displays have a memory of up to 100 hr. Additionally, the application of the field for the entire switching duration is not necessary. The pigment particles will move to the

opposite electrode on their own once they have cleared the potential wells in which they were stored; however, the switching speed is dependent on the applied voltage. Appxoximately 20 msec is required for a 50-V device to switch states. The characteristics for state-of-the-art electrophoretic displays are presented in Table 10.1.1.2.3-1.

As with any panel display with a large number of elements, the need for multiplexing exits. The electrophoretic display lacks the inherent nonlinearity necessary for multiplexed matrix addressing. The linear characteristics of the display will turn cells partially on when they are in the row or column being addressed. A control electrode concept has been developed to overcome this matrix addressing problem. The grid-type electrode adds a threshold to the cell, and since the threshold is more than half of the driving voltage, a single cell can be addressed without partially turning on other cells.

The electrophoretic cell provides good contrast provided the particle and the supporting fluid are made of highly contrasting colors. For example, black and white have been tried with excellent results.

A number of problem areas exist in the stability of the colloidal suspension, however. The pigment and suspending liquid must have exactly equal specific gravities so that the particles do not float or settle out and they must have electrostatic compatibility. Electrostatic compatibility means that the proper amount of repulsion must exist between the particles so that they do not have a tendency to form lumps. The pigment and liquid must also have a chemical stability so that they maintain their characteristics for an extended lifetime.

The size of electrophoretic display arrays built to date have been relatively small. Although no large displays have been built, results indicate that a matrix of 2,000 by 2,000 elements is possible with the switching characteristics of the control grid cell mentioned previously.

TABLE 10.1.1.2.3-1. CHARACTERISTICS FOR STATE-OF-THE-ART ELECTROPHORETIC DISPLAYS

CHARACTERISTIC	SPECIFICATION
Required Voltage (V)	15 to 50
Power Consumption ($\mu W/cm^2$)	Less than 5
Speed (msec)	20
Memory (hr)	100
Lifetime (hr)	8,500
Number of Operations (cycles)	Greater than 10^8
Temperature	
Operation (°C)	-10 to 70
Storage (°C)	-40 to 100
Appearance	Excellent contrast, viewing angle of a printed page
Color Availability	Many contrasting color combinations including black/white

10.1.1.2.4 Liquid Crystal Displays - The Liquid Crystal Display (LCD) is another type of passive display device that has gained widespread use. LCDs fall into two major categories: 1) the older dynamic scattering devices and 2) the newer twisted-nematic, or field-effect, devices. In both devices, an electric field applied across the liquid crystal material causes a change in the structure of the crystal with a resulting change in its effects on light. The field-effect LCD has replaced the dynamic scattering display in almost all applications. The field-effect device works on the principle of light polarization. A light polarization filter is placed both behind and in front of the LCD display. With no voltage applied to the cell, the crystal structure aligns to make the display appear transparent. When a voltage is applied to the cell, the crystal realigns to block light passage between the filters and the display cell appears dark.

The LCD has several advantages over other types of display media. The most noted of these is the low-power consumption, which is on the order of 5 μW per cell. This low power consumption is especially useful for displays designed for battery operation (such as wristwatches). Another advantage of the LCD is its contrast under conditions of high ambient light. An active display (such as the CRT or LED) will look "washed out" in bright sunlight but the LCD has a sharp contrast and is easy to read.

The field-effect cell has the disadvantage of requiring a polarizer, a component that has given poor reliability in the past. The most commonly used polarizer material is an iodine compound that is sensitive to both high humidity and high temperature. The humidity problem has been partially solved by sealing the polarizers in a laminated moisture-tight structure. An alternative solution is the use of the "K polarizer", manufactured by Polaroid Corporation. This is a more complex chemical that does not diffuse under humid conditions. Disadvantages of the K polarizer include high cost and the limited viewing angle of the field-effect cell. The viewing angle of the K polarizer is restricted typically to about 45 deg on either side of center.

One problem common to all LCDs is the time taken for turn-on and turn-off. At room temperature, the rise and decay times for field-effect displays are in the range of 200 msec. The rise and decay times increase at lower temperatures and can be over a second at 0°C. A new type of crystal, called phenylcyclohexane, has been developed for low-temperature operation. The rise time at 0°C is 380 msec and the decay time is 270 msec. Other materials have been developed that exhibit the liquid crystal properties from -55°C to +80°C, but the rise time at the low temperature is several hundred seconds. To overcome the temperature-related problems, several companies are developing LCD heaters. A thin-film heater is placed behind the display panel and is used to keep the liquid crystals operating with faster switching characteristics. Care must be taken with this method to avoid damage due to thermal cycling.

One of the requirements of any large-scale display is the need for efficient multiplexing techniques. LCDs must be driven with ac signals only, and this presents a problem for multiplex drivers since a symmetrical waveform must be switched. Additionally, the liquid crystal material does not exhibit nonlinearity in the voltage threshold curve and is subject to partial switching of display cells. The slow switching speed mentioned previously is also a problem when trying to obtain multiplexed operation. Some of these multiplexing problems are being solved by using crystal materials that have both faster rise times and a non-linear voltage threshold characteristic. Current crystal compositions allow the multiplexing of several digits, but large-scale multiplexing has not yet become practical for commercial applications.

Intensive research is being placed into the development of pleochroic dyes for liquid crystal displays. These displays show white numerials on a colored backgrour 1. They do not require polarizers, and they have a wide viewing angle of 180 deg. These displays require a higher drive voltage of 8 V, compared with 3 V for field-effect LCDs. The switching speeds are 120 msec rise time and 150 msec decay time at room temperature. Displays currently available exhibit white characters on either a purple, black, or blue background.

A recent development is the fluorescence-activated liquid crystal display (FLAD). This device has a fluorescent plate placed behind the liquid crystal plate. Ambient light stimulates the fluorescent panel and enhances the display. The FLAD can be read in minimal light while continuing to exhibit good contrast in conditions of high ambient light.

Several video-type displays have been constructed using liquid crystal technology. In early 1976, Hughes Aircraft demonstrated a display having 175 by 175 elements on a 1.75-in. square plate. In mid-1977, the Japanese released information on a black and white television with a 6-in.-diagonal screen. It has 16 shades of gray and a contrast ratio of 20:1. The response time of the liquid crystal is much longer than the TV line time, so the display is unable to keep up with fast motion without blurring. The display is driven by CMOS, and the total display power requirement is only 10 mW. Meanwhile, the French announced in late 1977 that they are developing a miniature LCD for video phone operation. The French display measures 6.4 by 6.4 mm and contains a 128 by 128 element matrix. The display has a center-to-center element spacing of 0.05 mm, and eight shades of gray are achieved with this display. The actual display is too small for direct viewing, so the image is projected with a lamp and lens arrangement.

10.1.1.2.5 Other Display Technologies - The display technologies covered in this paragraph are each important in their own respects, but the coverage given them is limited due to their limited value for terminal display applications. No attempt will be made to cover the many unique and novel ideas that are far from the commercial market.

Thin-film transistors (TFTs) were mentioned in conjunction with powder electroluminescent panels. TFTs themselves do not form a display, but they offer the potential as display drivers for many of the other display elements. TFTs have been used in conjunction with a number of display technologies including electroluminescent phosphors, liquid crystals, and gas discharge devices. The TFT enhances the capabilities of the other technologies by providing display storage capability and by providing the nonlinearity necessary for multiplexed operation of large matrix displays.

The TFT has a substrate made of glass and the dimensions can be several feet each way. The TFT driver panel can be made any size necessary to accommodate the display being driven; even large sizes are feasible. Furthermore, the TFT panel can be designed to accommodate 300-V switching requirements, so they are directly compatible with the high-voltage plasma displays. The display driver can be deposited on the same substrate as the display element for sparse displays, or it can be deposited behind the display element for dense display panels. In view of the flexibilities provided, it appears almost certain that TFTs will be used extensively in display systems of the future.

The Light Emitting Diode (LED) is a solid-state active display technology that has gained widespread use for small displays. The LED is a popular display device for watches and calculators and has received some use as an alphanumeric display panel where only one line is required. The cost of the display materials and the need for large peak currents when driving multielement displays limit the usefulness of the LED for TV-type graphic displays.

The LED is used in many forms for display devices. In addition to the standard LED "bulb", LEDs come in display blocks for 7- and 16-segment readouts and for 5- by 7-element dot matrix displays. In addition to the traditional red color, LEDs are available in green, yellow, and orange. Designers are working on devices with a blue radiation, but have had no success to date.

Despite the high cost and driving problems, many researchers are working on LED panel displays. Hewlett-Packard has developed a 30- by 36-element matrix on a 0.5- by 0.5-in. substrate. The yield for a perfect matrix is currently only about 5%, but if a matrix is allowed to have four defective elements and if a capacitive discharge repair system is used, the yield increases to 54%. Other work in this area includes a 14- by 15-element matrix developed in Japan. This device uses both red and green LEDs to give it a limited color capability. In addition, the LEDs are of a four-layer design that provides inherent memory.

Currently, indications are that it is unlikely that the LED matrix will ever approach a commercial-quality television display, since thousands of amperes of current would be required to drive the display.

Rotating balls are another display technology that is receiving some research work. These devices use very small balls and should not be confused with the large electromechanical displays. Two systems are currently under investigation for small-screen displays: magnetic balls and electrostatic balls.

The magnetic balls are microscopic spheres made of a ferrite material and painted white on one hemisphere. When an external magnetic field is applied, the spheres will rotate to show either white, black, or a shade of gray. The display is activated from a matrix of orthogonal current carrying wires behind the panel. The intersection of two of these wires produces enough magnetic field to cause the spheres to switch. Nonlinearity of the magnetic material provides a threshold, and the hysteresis of the ferrite provides a memory for image storage. Power consumption for this display is low provided that the information is infrequently changed.

The electrostatic display is likewise made of microscopic spheres with white on one side and black on the other. The composition of the sphere is a dielectric material. These balls respond to an electric field interacting with the fluid surrounding the spheres. The electrostatic display does not have a threshold, but it does exhibit a memory. Like the magnetic display, power consumption is low for infrequently changed data.

There are many other displays in use, such as the vacuum fluorescent and incandescent displays. Both of these have a market for applications with limited information content, but neither is suited for displaying more than a line or so of information. Other small-screen display technologies exist, but they do not possess the potential to be included in a technology assessment at this time.

10.1.2 Trends and Projected Developments in Small-Screen Displays

Small-screen displays as discussed herein include alphanumeric displays, graphic displays, and imaging systems. The trends and projected developments are presented first in terms of the available terminal features and second in terms of the display medium. As with other areas of data systems, the technology will be reflected largely by market demand and not necessarily by feasibility or availability.

Advances in the capabilities of available features in display devices can be expected to continue to increase at an accelerated pace. Further development of integrated circuit technology and microprocessors will give added capability and intelligence to the display terminal while lowering the power consumption and reducing the size. Advances in communications hardware and the implementation of distributed processing will give the user of remote display terminals more computing power with shorter turnaround time. The implementation of both large semiconductor memories and magnetic bubble memories will give added capabilities in data storage, in addition to adding display features. For more information on each of these related technologies, refer to Sections 3, 4, 7, and 9.

Recent advances that have resulted from the supporting technologies mentioned include the introduction and growth of the intelligent terminal, the replacement of analog circuits with digital circuitry for graphics manipulation, and the increased use of MOS memories for storing and refreshing relatively high-resolution color graphic displays, such as the Ramtek display system. The improvements in IC technology speed, packing density, power dissipation characteristics, and packaging that are forecast in previous sections (primarily Section 7) will be reflected on almost a one-for-one basis in future terminals. Features for alphanumeric display terminals of the 1980-1985 timeframe are shown in Table 10.1.2-1. The characteristics of graphic display systems for this period are shown in Table 10.1.2-2. The features of graphic

TABLE 10.1.2-1. PROJECTED DEVELOPMENTS IN ALPHANUMERIC DISPLAY TERMINAL FEATURES

FEATURE	PROJECTION
Memory	64 Kbytes standard >64 Kbytes common, 1 to 16 Mbytes available
Microprocessor Control	Generally used on most systems
Edit Features	
Cursor Positioning	U, D, L, R, H on all
Cursor Blinking	General availability
Character Insert/Delete	General availability
Line Insert/Delete	Wide availability
Erase	Character, line, screen
Roll	General availability
Paging	General availability
Color	4 colors generally available, blended multicolor limited availability
Auxiliary Storage	General availability
Light Pen/Other Select Modes	General availability
Selectable Brightness	Generally 2 levels
Compatibility with Standard Devices	Generally available
Internal Refresh	On most models

TABLE 10.1.2-2. PROJECTED DEVELOPMENTS IN GRAPHIC DISPLAY SYSTEM CHARACTERISTICS

CHARACTERISTIC	PROJECTIONS
Display Medium	CRT most common plasma on some systems, LCD and EL by 1985
Maximum Viewable Area	300 in^2
Addressable Matrix	8,192 × 8,192 typical, 64K × 64K available
Viewable Matrix	2,048 × 2,048 typical, 8,192 × 8,192 maximum
Word Length	16 bits
Points/Second	10^6 common, 5×10^6 available
Vectors/Second	10^6 in/sec common, 5×10^6 in/sec available
Character/Second	3.3×10^5 common, 10^6 available
Character Set	128-character ASCII set typical Additional symbols and languages common

display systems are given in Table 10.1.2-3. The alphanumeric terminals will be capable of incorporating increased processing capabilities and will have both random-access and high-speed auxiliary storage, probably in the form of magnetic bubbles. In this timeframe, terminals with the processing capabilities of today's minicomputer will be available at costs that rival a medium-capability intelligent terminal today. Random-access storage of 64 Kbytes should be standard, with capabilities to 1 Mbyte being common. (Technology will support higher capacities, but the availability will be a function of market demand.) Bubble memories with capacities of 1 to 16 Mbytes can be expected in a few systems. Internal refresh (via high-speed semiconductor memory) of high-resolution (2,048 by 2,048 pixels) color graphic displays and/or images will be feasible in terms of both cost and performance. The added processing and/or storage capacity in display devices will provide almost all required processing capabilities within the graphics or image processing terminal, thus leaving the host processor free to perform more important functions.

In terms of the display medium, the CRT will continue to be the dominant device in the 1980-1985 timeframe. The CRT will continue to exhibit its present characteristics with only minor changes expected. Variations of the CRT (such as the flat CRT) will be available but will not be in mass production and will continue to be costly. Plasma devices will become more economical and will be used for specific applications requiring a display with mechanical ruggedness and a thin profile. Electroluminescent displays will become available on a limited basis by 1980 and will be capable of good-quality video display by 1985. The widespread use of EL panels for television will not come until the late 1980's, however. Electrophoretic displays will be used for smaller displays not requiring matrix operation or high-speed response. The liquid crystal display will become one of the major competitors of the CRT, at least for devices not displaying motion. The slowness of the liquid crystal response will limit its usefulness for standard television,

TABLE 10.1.2-3. PROJECTED DEVELOPMENTS IN GRAPHIC DISPLAY SYSTEM FEATURES

FEATURE	PROJECTION
Color	4 colors generally available, blended multicolor available on some
Intensity/Shades of Gray	General availability
Perspective (3-D) Views	Available on some
Translation/Rotation/Offset	General availability
Line Texturing	General availability
Windowing/Zooming	General availability
Conics	General availability
Refresh Mode	Widely available
Internal Processor	Widely available
Source Languages	FORTRAN, COBAL, BASIC, and high-level problem/system oriented languages

but it will prove well-suited for computer graphics displays. Widespread use will not come until the mid-1980's, however. Other display technologies will continue to develop, but none is expected to have great impact on computer-oriented displays. Table 10.1.2-4 presents the projected developments for the various display media.

TABLE 10.1.2-4. PROJECTED DEVELOPMENTS IN DISPLAY MEDIA

DISPLAY TECHNOLOGY	PRIMARY INCREASE IN CAPABILITY	PRIMARY APPLICATION
CRT	Slight increase in resolution	Will remain the primary display medium for video and computer graphics
Plasma Panel	Lower cost	Will be used for specific applications requiring mechanical ruggedness and thin profile
Electroluminescent Panels	Longer life, improved color gamut	Alphanumeric display in early 1980's, video display by 1985
Electrophoretic Panels	Longer life, improved nonlinearity	Small displays not requiring high-speed response
Liquid Crystal Panels	Improved life, improved matrix capability	Alphanumeric displays, video displays not requiring motion, computer graphics displays
Other Technologies	Various	Small alphanumeric displays

10.1.3 State of the Art in Large-Screen Displays

Large-screen displays as discussed herein represent those displays that are greater than 25 in. in diagonal measurement. These displays fall into the general categories of matrix systems and projection systems. The matrix systems use either active arrays of lights or an array of passive characters. The projection systems use either direct magnification of a CRT image or use a light valve and projection bulb.

The more common form of large-screen matrix display uses an array of lights that can be controlled to form both alphanumeric and image displays. This type of display has widespread applications for information and advertising signs. The major improvement to these displays over the past few years has been the replacement of the mechanical sequencer by solid-state control electronics. Large arrays of this type have been implemented for image display in stadium environments. A typical array of this type contains nearly 10,000 pixels, with the picture controlled by a microprocessor. Displays of this type have the disadvantage of high power requirements, with some of the larger displays consuming hundreds of kilowatts of power.

The passive electromagnetic displays offer capability for alphanumerics, symbols, and limited image display. These displays are available with characters ranging in size from 1 to 24 in. The principal applications are variable message signs such as are used for airport announcement boards, stock exchanges, and outdoor advertising. The display elements consist of a disk and magnet mounted on a common axis. An electromagnet nearby can be switched so that the disk will rotate to show either a light or dark image. The display exhibits memory, and power is needed only when the information is being changed. A pulse of 3 A for 300-μsec duration is sufficient for changing a single display element. By updating the characters in sequence, the peak current can be kept low. Changing the display at a rate of 10 characters per second

consumes approximately 33 W. Since the display is passive, lighting must be added for night viewing. In this case, light is projected onto the display much as on a highway billboard. The power consumption for a 100-character display with night lighting is approximately 100 times lower than for an equivalent lamp matrix display.

Projection-type displays offer more capability for both alphanumeric and image displays than can be found in matrix displays. The simplest of the projection displays uses a lens arrangement to project the image from a CRT directly onto a screen. This is the technique used in the home video projection systems, and while it is the least expensive method, the quality is somewhat lacking. The limit on available light from the CRT display produces an enlarged image that is dim. The raster of the CRT also creates problems as the enlarged raster emphasizes the low resolution and magnifies any nonlinearities in the video display.

The commercial-quality projectors use a light valve and a projection bulb to increase the available light on the display. The image can be written on the light valve with a CRT, an electron beam, or a laser beam. The composition of the light valves varies from company to company. Several systems are available on the commercial market, and several others are in the laboratory development stage. Table 10.1.3-1 presents the state-of-the-art characteristics for light valve projection systems.

The most common type of light valve system in current production uses an oil film to modulate the light from a projection lamp. Both the Eidophor and the General Electric systems use this technique. The Eidophor color television projection system uses an electron beam to scan an oil-covered surface on the light valve. A xenon projection lamp is reflected from the light valve and aimed through a lens to the projection screen. The General Electric system likewise uses an electron beam to deform an oil layer on the light valve. This system uses a

TABLE 10.1.3-1. STATE OF THE ART IN LIGHT VALVE PROJECTION SYSTEMS

MANUFACTURER/DEVICE	TRANSMISSION MODE	LIGHT MODULATION MEDIUM	WRITING SCHEME	EFFICIENCY (lm/W)	REAL-TIME VIDEO CAPABILITY
Advent/Cathodochromic CRT	Reflective	Sodalite	Electron beam	-	No
Conrac/Eidophor	Reflective	Oil film	Electron beam	1.5	Yes
General Electric/ GE Light Valve	Transmissive	Oil film	Electron beam	1.2 to 1.5	Yes
Hughes Aircraft/*	Reflective	Liquid crystal	Light beam	-	No
IBM/*	Reflective	Liquid crystal	IR laser	-	No
Xerox/Ruticon	Reflective	Metallized elastomer	Light beam	-	YES**
Western Electric	Transmissive	Liquid crystal	IR laser	-	No

*Liquid crystal projection system.
**With degradation of upper 18% of picture.

transmissive light valve, however. By using polarized color filters, the GE system is able to achieve simultaneous R.G.B. color with a single electron gun and a single projection lamp. Both of these systems have response times compatible with commercial television and can be used for real-time video display without blurring or loss of motion.

Another type of light valve developed by Xerox, and called the Ruticon light valve, uses a deformable metallized elastomer as the reflecting surface. The light valve is exposed from the rear using the light from a CRT. This light falls upon a photoconductive surface that varies an electric potential across the cell. This deforms the reflective elastomer surface, and light striking the front of the valve is reflected to match the image from the CRT. The Ruticon has an inherent memory and must be shorted and exposed to an intense light in order to erase the memory. This presents a problem for real-time video display since the valve must remain shorted for a period of time much longer than the video field retrace of the CRT. When used in this mode, the upper 18% of the video image lacks resolution and appears dark because the tube is shorted during the beginning of the first field scan. The second field scan does write on the upper part of the reflecting surface, but resolution and light intensity are lowered by 50%.

Another reflective light valve is the cathodochromic CRT (CCRT), which employs sodalite as the cathodochromic material. The sodalite target is scanned with an electron beam to make alternate areas either reflective or transmissive. A rear projection light is beamed onto the target and the resulting image magnified and projected. The image is erased by writing on the entire target with an electron beam. The erase cycle at present takes 2.8 sec, so the system does not have the response necessary for real-time video display. The CCRT display is suitable for high quality alphanumeric display applications. This type of display has a writing speed of approximately 580 characters per second in the black-on-white mode and 1,060 characters per second in the white-on-black format.

Several systems have been developed for graphic and alphanumeric display using liquid crystal light valves. Hughes Aircraft has announced a liquid crystal projection system with 1,024-line resolution. Their liquid crystal light valve consists of a sandwich of liquid crystal material and photoconductive material between two transparent electrodes. A bias voltage is applied across the cell and light striking the photoconductive layer causes the liquid crystals to be turned on. Polarized light from the projection lamp illuminates the polarized liquid crystal light valve and the polarization of the crystals determines the light that is reflected or absorbed. The reflected light is projected onto the display screen. The slow speed of the liquid crystals limits the use of this technique to alphanumeric and graphic display applications.

Western Electric has developed a system that uses an infrared laser to create hot spots and thus write onto a liquid crystal light valve. The system was developed for creating printed-circuit artwork but has potential for graphic and alphanumeric projection. The laser beam is moved across the light valve using galvanometer-controlled mirrors. An arc lamp behind the liquid crystal valve provides the light for projection. The system has a capability of writing over 147,000 characters at a speed of 1,430 characters per second using a 7 by 9 dot matrix.

A similar system designed by IBM uses an uncooled semiconductor laser to write on the liquid crystal light valve. A pair of x-y galvanometer mirrors are used to control the laser beam. Reflective aluminum is deposited on the rear of the valve, and crystal polarization determines whether the cell reflects the light from the projection bulb. Present capabilities allow writing at a rate of 20 characters per second, with a total display of 2,000 characters available.

10.1.4 Trends and Projected Developments in Large-Screen Displays

The major improvements during the next few years for large-screen displays will be in the projection systems rather than in the matrixed systems. The matrix-type devices will show no dramatic improvement during the next several years; however, some improvements will be made in the control and drive electronics. Very little improvement is expected in the actual display elements for large-screen displays. The improvements in projection-type systems will be dependent on developments in light valve technology. The oil layer valves currently have the best capability for real-time video and will probably continue to lead in this area. The image quality should continue to improve for the other projection display tubes, with the efficiency increasing for all systems.

The liquid crystal projection systems offer the greatest potential for graphic and alphanumeric large-screen displays. These systems provide the sharpest contrast and best resolution of any of the projection type systems. Progress with liquid crystal projection systems will be linked mainly to two technology areas: 1) improvements in liquid crystal materials and 2) improvements in semiconductor laser diodes. The entire display industry is striving to improve the liquid crystal display, and the communications industry is enhancing the laser diode for use in fiber-optic transmission links. Liquid crystal projection display technology will therefore benefit both from direct research in liquid crystal display systems and from research in the associated fields.

10.2 HARDCOPY INFORMATION PRESENTATION ELEMENTS

The primary media used for hardcopy information presentation are the photographic film and the paper page. The photographic data presentation elements discussed in this report are computer output microfilm (COM) and microfiche. The paper hardcopy is addressed in terms of the printers and plotters required to produce the page; specifically, the electro-mechanical printers and plotters (as opposed to the mechanical printing presses). The discussion of printers includes low-speed serial devices; however, the emphasis is placed on the medium-to-high-speed devices used for computer output. The discussion of plotters covers both the small hardcopy devices used for time-shared terminals and the larger plotters used for output from both minicomputers and large mainframe systems.

10.2.1 State of the Art in Computer Output Microfilm

Computer output microfilm (COM) systems are available as both on-line and off-line equipment, with the off-line systems normally using magnetic tape recorders for input. These systems are capable of producing alphanumeric and graphic data at rates to 30,000 lines/min for alphanumerics and 400,000 points/sec for graphics.

Available COM devices may be configured in many different ways. Some manufacturers provide interfaces to only the IBM 360/370 series; others provide interfaces to any third-generation computer; and still others claim compatibility with any computer. Many of the newer systems are interfaced via front-end processors that increase the flexibility but may require user-provided software. Available features on COM equipment vary widely, ranging from multiple fonts, the presence or absence of graphic capability, different reduction ratios, different film sizes, and number of characters per line. A number of these features are presented in Table 10.2.1-1, which gives a representative listing for typical computer output microfilm systems. The details and implications of these features are discussed in subsequent paragraphs.

The majority of the COM equipment on the market today uses a CRT to expose the film. As the technology improves in dynamic displays, different types of image-generation systems are being employed. Two of the more innovative systems use lasers and light-emitting diodes. The 3M Company uses both an Electron Beam Recorder and a Laser Beam Recorder to write data directly onto the film. Memorex composes the output onto a bank of light-emitting diodes (LEDs) through a character-translation matrix. The light from these diodes is transmitted to the film through fiber-optic strands to form a character image. One difficulty in this latter approach is that the image is generated a line at a time (CRT imaging systems generate a frame before the film is advanced, although the film may be exposed character-by-character). This imaging method has limited capabilities for producing graphics

TABLE 10.2.1-1. STATE OF THE ART FOR COMPUTER OUTPUT MICROFILM SYSTEMS

MANUFACTURER	MODEL	PRINT SPEED	CHARACTER SET	CHARACTERS PER LINE	INPUT	FILM FORMAT	REDUCTION RATIO
Datagraphix	4530	14,000 lpm 60,000 cps	169	80 132 144 160	7/9 track mag tape 800/1,600 bpi	105-mm	48× 42× 24×
Datagraphix	4560	20,000 lpm	190	80 132 144 160	7/9 track mag tape 800/1,600 bpi	std: 105-mm opt: 16-mm 35-mm 82.5-mm	48× 42× 24×
Information International	8032	40,000 cps	-	-	9 track mag tape 800 bpi	105-mm	48× 42×
Kodak	KOM-80	-	std: 64 opt: 82	132	7/9 track mag tape 800/1,600 bpi	105-mm 82.5-mm 35-mm 16-mm	48× 42× 24×
Kodak	KOM-85	120,000 cps 30,000 lpm	std: 64 opt: 121	std: 132 opt: 172	7/9 track mag tape 800/1,600 bpi	105-mm 35-mm 16-mm	std: 48× 42× 24× opt: 32×

because of the difficulties in controlling movement of the film. Use of the LED approach also restricts the generation of microfiche because of inability to control the film. This latter difficulty may be a greater limitation on the use of this technique than the limited graphics capability because more users appear to be using microfiche than microfilm in today's market. One major advantage of LEDs is that they produce up to a thousand times more light energy than a CRT and thereby lessen the effects of film sensitivity. Thus it appears likely that attempts will be made to overcome the problems identified above to enable the use of this technique on future COM devices.

Reductions in size of 42× to 48× appear to represent the state of the art. Some vendors are unable to overcome the resolution problem in going to 48× reductions. Silver halide films typically have resolutions that range from 100 to 190 lines per millimeter, and films with resolutions of 190 lines per millimeter or higher are required to accomplish reductions of 48:1. Silver halide film is considered the industry standard for the recorder master copy. One of the highest-grade films available for COM applications is Kodak AHU, and the price is reflected accordingly. The most widely used film for duplicating is designated Diazo. Diazo is sensitized with diazonium salts and exposed by strong ultraviolet light. Ammonia is used in the development process to yield a direct image. Diazo is less expensive to purchase and to process than other types of COM film. A relatively recent introduction is vesicular film, which has a light-sensitive coating on a plastic backing. The film is exposed by ultraviolet radiation and is developed by the application of heat. It is replacing Diazo in many installations because of the simpler processing chemistry that eliminates the need for ammonia.

As mentioned above, the introduction of front-end processors on many COM systems has increased the flexibility of these systems. These processors not only relieve the host processor of its processing load and provide DMA-to-DMA interfaces to speed up the data throughput, they

also provide capabilities for handling standard print tapes that reduce the operator burden and/or the need for operating system modifications within the host processor. These front-end processors perform a number of functions that otherwise have to be performed by the host processor or deleted. Among current functions being performed are formatting, graphics generation, control functions, character selection, software translation, and film indexing. These functions can be expected to increase as the costs and capabilities of such processors continue to decline.

10.2.2 Trends and Projected Developments in Computer Output Microfilm

Computer output microfilm is another area of technology that has failed to grow as rapidly as many experts projected in the 1960's. The probable cause is a prejudice against microfilm. Many potential users are hesitant to eliminate the paper copy. Additional disadvantages are the special reading and filing equipment, the lack of direct update capability, and the inability to write comments directly on the page. In spite of the lack of public acceptance, the technology has followed a relatively steady increase in capability. The most important recent improvement in COM systems is the introduction of the front-end processor to increase system flexibility. Front-end processors will be used more and more to enhance the capability and flexibility of COM equipment. Also, advances in the area of image-generation devices will continue to be reflected in the COM systems of the future.

Datapro Research Corporation (Ref. 10-3) feels that COM prices will not decline in the foreseeable future. Instead, any reduction in component prices will be reflected in greater sophistication and flexibility.

10.2.3 State of the Art in Printers and Plotters

The state of the art in printers and plotters is presented in terms of impact printers, nonimpact printers, and plotters. The discussion of each type includes a brief discription of the methods currently used, followed by their characteristics and capabilities.

Printer technology covers a wide range of capabilities and applications. Typical impact printers range from 10 characters per second (cps) to 2,250 lines per minute (lpm), with the cost of these devices varying from less than $1,000 to $112,000. Nonimpact printers range in capability from 30 cps up to 45,000 lpm. Costs for these printers range from less than a thousand dollars for several electrographic printers to well over $300K for the very-high-speed laser xerographic and ink-jet printers. Applications for these printers range from cash register printouts and small status monitoring equipment to high-speed computer output. Current estimates include an availability of some 400 models from some 100 manufacturers.

10.2.3.1 Impact Printers - Impact printers by definition employ some type of mechanical force to place the character image onto the paper. The design of the print mechanism can be further defined as either a template or a matrix. The template printers use some form of pre-engraved type, with the reproduced characters having the appearance of the type. The matrix printers use a number of pins arranged so that they impact to form a matrix of dots, with the characters being determined by the dots reproduced or omitted. The methods used to implement impact printers vary, as do the speed and cost of the devices. One important characteristic shared by all impact printers is the capability to produce multiple copies at once. This can be accomplished using multiple ribbons and/or carbon packs or by using pressure-sensitive paper. The ribbon and carbon pack impact printers also have the advantage of using plain paper, which is a feature not found in many of the nonimpact printers.

Table 10.2.3.1-1 lists the major characteristics for typical state-of-the-art impact printers. The simplest and most inexpensive

TABLE 10.2.3.1-1. CHARACTERISTICS FOR STATE-OF-THE-ART IMPACT PRINTERS

PRINTER TYPE	SPEED	ADVANTAGES	DISADVANTAGES	COST ($)	MAIN SUPPLIER
Cylinder	10 cps	Availability Low cost	Low Speed Noisy operation Poor print quality	1 to 2K	Teletype Corp.
Golf Ball	15 cps	Good print quality Interchangeable fonts Relatively low cost	Low speed Noisy operation	2 to 5K	IBM
Daisy Wheel	30 to 55 cps	Higher speed than golf ball or cylinder Interchangeable fonts	Print quality not as good as golf ball	3 to 7K	Diablo Systems, Inc. (Xerox)
Serial Matrix	30 to 330 cps	Fine resolution on some Interchangeable fonts Some graphics capability Higher speeds on some	Dot characters can be difficult to read Reliability drops with high duty cycle	2 to 4K	Centronics Data Computer Corp.
Line Matrix	500 lpm	Fine resolution on some Interchangeable fonts Some graphics capability Relatively high speed	Dot characters can be difficult to read Reliability drops with high duty cycle	5 to 9K	Tally Corp.
Drum	300 to 2,000 lpm	Reliable operation High print speed Good performance at medium to heavy duty cycle	Limited choice of fonts Vertical misregistration of characters	10 to 70K	Data Products Corp.

TABLE 10.2.3.1-1 - Concluded

PRINTER TYPE	SPEED	ADVANTAGES	DISADVANTAGES	COST ($)	MAIN SUPPLIER
Train (Chain)	2,250 lpm	Good print quality Variable character sets, up to 128 characters	Not as reliable as drum	10 to 112K	IBM
Band and Belt	300 cps to 2,000 lpm	Interchangeable fonts Good print quality High Reliability	Subject to belt wear Poor print quality with some types Entire belt must be replaced if character becomes defective	3 to 87K	General Electric IBM

impact printing technique is the cylinder printer. This system is used on the familiar teletype machines. The cylinder printer uses a hammer to strike a cylinder upon which characters are embossed, and the characters are printed through a ribbon onto the paper. The cost of these devices is good, but they are plagued by low speed (10 cps), noisy operation, and poor print quality.

Another printer type of which is similar to the cylinder is the "golf-ball" printer. This device functions like the cylinder except that the characters are embossed on the outside of a ball-shaped print head. The golf-ball printer offers slightly greater speed (15 cps) and has the advantages of good print quality and interchangeable fonts. These devices are inherently noisy, however, and are still slow for most applications.

A third type of similar printer is the "daisy wheel", in which a single character is embossed on a spoke and a number of spokes are located around a central hub. The appropriate spoke is then tapped with a hammer to imprint the symbol through a ribbon. The daisy wheel can operate at speeds of 30 to 55 cps and has easily interchangeable fonts. The print quality, however, is not as good as the golf ball.

In terms of the larger and faster template impact printers, the drum printer is the best known and the most reliable. The drum has a complete character set embossed around its circumference for each print column. The drum rotates at a constant speed and a hammer strikes the paper and ribbon against the appropriate character. The print speed for the drum printer ranges from 300 to 2,000 lpm. It offers reliable operation and good performance for medium- to heavy-duty use. There is a limited choice of character fonts, however, and the problem of vertical misregistration causes wavy lines due to some characters being above and below the line.

Another type of high-speed printer is the train or chain printer, which incorporates a set of characters moving horizontally across the page, with each character in the form of a slug. The slugs are connected

and pull each other around in the chain printer. The slugs are not connected in the train printer, and each piece pushes the next around a track. A hammer in each print position strikes the paper against the type as the appropriate character is aligned. Typical printers can accommodate up to 128 characters in a set. The speed of these devices runs up to 2,250 lpm. The advantages are good print quality and variable character sets, but they have a lower reliability than the drum printers. An advantage of the chain printer is that any misregistration in this printer is in the horizontal direction and is not as noticeable as the vertical misregistration of the drum printer.

The band and belt printers are somewhat similar to the chain printers. The characters are embossed in a continuous steel belt that moves across the paper in a horizontal direction. A hammer sits at each character position and strikes the paper against the type as the appropriate symbol passes. The print speed ranges from 300 cps to 2,000 lpm. The advantages of the band/belt printers include high reliability, interchangeable fonts, and good print quality. Character misregistration is in the horizontal direction as in the chain-type printers and is not very noticeable. The belts are subject to wear, however, and the entire belt must be replaced if a single character becomes defective.

Matrix impact printers form the desired character from a pattern of small dots. The dots are formed when small needle-like hammers are struck against the paper. The number of dots in the matrix varies among the different manufacturers, with typical configurations being 5 by 7, 5 by 9, 7 by 7, and 7 by 9. The 5 by 7 matrix machines are capable of forming 64 standard ASCII characters, while the 7 by 9 matrix can form both upper- and lower-case alphanumerics. The character codes are stored in a read-only memory, and the list of available symbols is limited only by the code storage space. Matrix printer speed, unlike many of the other types of impact printers, is independent of the number of characters in the set. The versatility of the matrix allows the printing of special symbols and various alphabets, as well as changing the appearance of the standard alphabet.

Mechanical impact matrix printers generally fall into two categories: serial-matrix and line-matrix. The serial-matrix printer has a single head that moves horizontally across the page. The vertical dots of each character are formed simultaneously, with the head scanning to produce the horizontal dots. In this manner, one scan across the page will yield one line of print. Serial-matrix machines are capable of forming 30 to 330 cps. They offer moderate to fine resolution, easily interchangeable character fonts, and a certain degree of graphics capability. The dot characters become difficult to read on the smaller matrix sizes, however, and the reliability of the printer drops as the duty cycle increases. To achieve a greater print speed, several of the serial printers use a bi-directional print scheme. The printer head moves from left to right, printing the desired message, and then prints from right to left on the following line to save the time taken for the head to return to the left margin. A line of message is loaded into the print controller and the reversing of the character sequence is automatically accomplished for the right-to-left print scan.

The line-matrix printers are capable of producing output rates of up to 500 lpm. The print head consists of a horizontal row of print needles contained on a bar or comb. The bar slides from left to right, with the needles striking each selected dot to produce a row. The complete characters are formed as the paper is moved to successive rows. The vertical resolution of a typical line-matrix printer varies from 70 to 100 dots per inch. The line-matrix printer offers fine resolutions with the larger matrix sizes, as well as easily interchanged fonts and relatively high speed. However, alphanumeric characters can be difficult to read when printed with the smaller matrix arrays. Table 10.2.3.1-2 shows a representative list of impact printers and their manufacturers.

10.2.3.2 Nonimpact Printers - Nonimpact printers use some form of indirectly placing the image onto the paper. They generally use special sensitized paper that can respond to electrical or thermal stimuli, as is done in thermal, electrographic, and electrostatic printers. Two developing

TABLE 10.2.3.1-2. TYPICAL IMPACT PRINTERS

MANUFACTURER	MODEL	PRINTING METHOD	SPEED	COST ($)	COMMENTS
Anadex	DP-1000	Matrix	50 cps	700	
Centronics	761	Matrix	50 cps	-	Teleprinter
Dataproducts	M-200	Matrix	200 lpm	-	
Dataroyal	7000	Matrix	120 cps	1,595	
Diablo	1660	Matrix	200 cps	-	Terminal
Florida Data Corp.	PB-600	Matrix	240 lpm	-	
Hewlett-Packard	2631A	Matrix	180 cps	-	
Lear Siegler	200	Ballistic Matrix	180 cps	-	
Practical Automation	DMPT-3	Matrix	120 cps	452	
Printronix	300	Matrix	300 lpm	-	
Sheldon-Sodeco	-	Spanning Hammer	90 lpm	140	21 Column, No Electronics
Teletype Corp.	40	Revolving Carrier	300 lpm	2K	
Dataproducts	2550	Band	2,000 lpm	-	
General Electric	-	Band	30 cps	3K	
Epson America	10	Belt	150 lpm	-	
IBM	-	Belt	2,000 lpm	-	
Dataproducts	2470	Drum	2,000 lpm	-	
IBM	-	Train	2,250 lpm	-	

technologies of nonimpact printers not using sensitized paper are the ink-jet printer and the laser xerographic printers. Both of these latter methods are especially suited for very-high-speed output devices, but the costs are quite high. Table 10.2.3.2-1 lists the characteristics for the state-of-the-art nonimpact printers.

One of the typical nonimpact printers is the thermal-matrix printer. The thermal-matrix printer typically has a print head with a 5 by 7 array of dot elements. The print head moves horizontally across the paper with the head stopping at each column. A recently introduced alternative method is to use a vertical column of seven dot elements that scan the page much like the serial impact matrix printer. In either method, the elements of the head are selectively heated. The leading type of print head is a 35-dot, 5- by 7-element array of single-crystal silicon mesas. Energy consumption for the silicon dots is 8 mW-sec/dot. Other thermal print heads use a resistive film in which a film resistor is deposited on a ceramic substrate. These print heads require 16 to 24 mW-sec/dot. Generally, thermal-matrix printers are among the least expensive and offer low-noise operation. They are ideally suited for small system applications, requiring 30 to 100 cps. This is relatively slow for many applications, and like all other nonimpact printers, no multiple copy capability is available. Additionally, the thermally sensitized paper is expensive and does not keep well in environments of bright light or high heat. These printers also lack the capacity for using pre-printed forms.

The electrographic printers use special paper that is sensitive to electric currents. A stylus print head moves across the paper and a current is applied to the stylus to form the characters. Two types of paper are commonly used. The electrolytic version uses a wet process in which moist paper is drawn between electrolytes. The application of a current to the paper then causes a chemical coloration. The electro-sensitive version uses dry paper with layers. As current flows from the stylus to the paper, the top layer of the paper is burned away, exposing a darker layer underneath. These printers are generally the lowest in cost, with prices starting below a thousand dollars. Additionally, the

TABLE 10.2.3.2-1. CHARACTERISTICS FOR STATE-OF-THE-ART NONIMPACT PRINTERS

PRINTER TYPE	SPEED	ADVANTAGES	DISADVANTAGES	COST ($)	MAIN SUPPLIER
Thermal-Matrix	30 to 100 cps	Low cost Low-noise operation Ideal for small operations	Relatively slow speed No multiple copy capability Paper is expensive No pre-printed forms	1 to 5K	Texas Instruments, Inc.
Electrographic	160 to 2,200 cps	Very low cost High serial print speed	Paper is expensive No multiple copy capability Paper looks and feels like aluminum foil	600 to 3K	Scope Data, Inc. SCI Systems, Inc.
Direct Electrostatic	300 to 18,000 lpm	Accommodates plotting Easily changeable fonts Lower cost than impact printer of same speed Very high speed possible	No multiple copy capability Paper is expensive Wet toner must be replenished	5 to 13K (Low speed) 165K (High speed)	Versatec, Inc. Honeywell Information Systems, Inc.
Laser Xerographic	4,000 to 21,000 lpm	Very high character resolution Quiet operation Multiple character fonts Accommodates 132 columns on 8.5 by 11-in. paper	No multiple copy capability High level of maintenance Limited to high-volume applications	145 to 310K	Xerox Corp. IBM
Ink-Jet	72 cps to 45,000 lpm	Character flexibility Quiet operation High print quality at lower speeds Very high speed possible	Poor copy quality at extremely high speeds No multiple copy capability	5 to 25K 5.8K/mo. rental	IBM Mead Digital Systems, Inc.

print speed is high for a serial printer, with speeds ranging from 160 to 2,200 cps. Disadvantages are mostly related to the paper. The paper has the look and feel of aluminum foil -- and wrinkles easily. The paper is also expensive and does not have the capability for multiple copies.

The direct electrostatic printer is capable of achieving high line speeds. In operation, a specially coated paper is passed over a row of fine metal styli. Each stylus is selectively charged to produce a group of charged spots corresponding to the character to be formed. The paper passes through a toner bath where ink particles are attracted to the charged areas. Some models use a scanning stylus instead of a fixed row. The scanning stylus moves across the paper, with the voltage being varied to produce charged areas of the desired character patterns. The paper is then run through the toner. The low-speed models use the scanning stylus and operate at speeds of 300 lpm. The fixed stylus design has no moving parts except for the paper transport mechanism. Therefore, very high print speeds of up to 18,000 lpm can be obtained. In addition to the very high print speed, a 2000-lpm electrostatic printer is lower in cost than an impact printer of the same speed. Another advantage of the electrostatic printer is that the fonts are easily changed simply by reprogramming the controller. The direct electrostatic printers also can be used for high-resolution graphics applications. Disadvantages of the electrostatic printers include the lack of multiple copy capability, the cost of the paper, and the need to replenish the wet toner.

One of the newer printer technologies is the laser xerographic printer. This printer uses a laser beam to scan a charged photoconductive surface and develop an image of charged dots. Wherever the laser beam strikes the photoconductive surface, the surface is discharged to leave a pattern corresponding to the image to be printed. The photoconductive surface is provided either on a drum or on a belt, as in conventional Xerox copiers. The image pattern picks up a dry toner and transfers it onto the paper. The resulting image is then heated to fuse the toner and render the image permanent. A light discharges the photoconductive

surface and a brush removes the excess toner from the belt or drum. The throughput of these printers is extremely high, ranging from 4,000 to 21,000 lpm. The character resolution of the system is extremely good, with one system using an 18 by 24 matrix to generate characters. The printer is quiet in operation since the paper drive is the major moving part. Multiple character fonts are available with 128 characters standard and an additional 127 characters optional. Because of the high resolution, smaller characters can be generated so it is feasible to place 132 columns on paper that is 8.5 in. wide. To date, the laser xerographic printing systems have been plagued by maintenance problems, but the technology is relatively new and the manufacturers feel that reliability will increase with experience.

Another developing printer technology uses the ink jet. This printer uses a stream of controlled ink droplets directed to the paper to form the desired characters. One of the schemes currently in use employs a horizontal bar with one hole per dot for the ink to pass through. A second design uses an on-demand ink-jet pump to deliver ink to a dot matrix. In a third design, an electric field deflects the droplets of ink to the appropriate places to form the characters.

The print speed for the ink-jet printers ranges from 72 cps for the office types to 45,000 lpm for the large computer output printers. Some other advantages of ink-jet printing are quiet operation, character font flexibility, high print quality for the lower-speed models, and extremely high throughput speeds for some models. Disadvantages include a high level of maintenance for the high-speed types, poor character quality at the high speeds, and the lack of multiple-copy capability. The use of ink-jet printers is increasing, however, and many of the current problems are likely to be solved as the technology continues to evolve.

Table 10.2.3.2-2 presents a representative list of nonimpact printers and their manufacturers.

TABLE 10.2.3.2-2. TYPICAL NONIMPACT PRINTERS

MANUFACTURER	MODEL	PRINTING METHOD	SPEED	COST ($)	COMMENTS
Perkin-Elmer	650	Thermal	20 sec/page	795	Terminal hardcopy
Texas Instruments	734	Thermal	30 cps	1,088	
Centronics	P-1	Electrographic	180 lpm	595	Small business
SCI Systems	-	Electrographic	2,200 cps	300	Highest speed/$
Scope Data, Inc.	1200	Electrographic	300 cps	-	Terminal printer
Versatec, Inc.	1640	Electrostatic	10 sec/page	10K	Terminal hardcopy
IBM	-	Xerographic	19,400 lpm	310K	
Siemens	-	Xerographic	21,000 lpm	-	
Xerox	-	Xerographic	18,000 lpm	-	
A.B.Dick	-	Ink jet	-	-	For marking containers
IBM	6640	Ink jet	72 cps	5K	Office machine
Mead Digital Systems	-	Ink jet	45,000 lpm	5.8K/mo.	Rental only
Siemens	-	Ink jet	300 cps	-	Office machine

10.2.3.3 Plotters - Plotters, unlike printers, are usually available as complete systems which include internal programmable controllers with the necessary software to produce complex plots with minimum supervision by the host computer. Plotters generally fall into one of three categories: flatbed, drum, and electrostatic. The flatbed plotter resembles an electromechanical drafting table in which the controller determines the pen position in the X and Y directions. The paper is held in a flat position by either electrostatics or vacuum. In the drum plotter, the pen moves along the X-axis of the paper. The paper is attached to the drum and the drum rotates to move the paper for plotting in the Y direction. The electrostatic plotter moves the paper across a row of styli and then runs it through a liquid toner. This is similar to the system used for electrostatic printers.

Table 10.2.3.3-1 presents a representative listing of state-of-the-art plotters. The majority of the plotters are either drum or flatbed and use pens for drawing the lines. A couple of the listed plotters use ink-jet techniques. Electrostatic plotters are becoming more popular, but they have the disadvantages of expensive paper and liquid toner. All of the entries are for complete plotter systems except for the device by Phillips, which is only the mechanism and does not include the controller.

The advancement of plotters took a leap in 1977 as several manufacturers introduced machines with microprocessor controllers. These machines reduce the amount of time taken from the CPU and reduce the time required to send the data from the computer to the plotter over a remote line. The data required to produce a plot are reduced by a factor of 20:1 in some cases, greatly reducing remote line charges. Additionally, the controllers are optimized for plotter control and can run several times faster than a minicomputer can supply the data. In addition to offering faster plotting, some estimates say that two-thirds of the work formerly done in the host computer can now be done by the plotter. In many cases, the vectors are the only information needed for the plotter to function.

TABLE 10.2.3.3-1. STATE OF THE ART FOR PLOTTERS

MANUFACTURER	MODEL	PLOTTING METHOD	PAPER SIZE (in.)	COST ($)	COMMENTS
Applicon	-	Ink jet	22 by 34	-	Drum, 3 colors
Gould Inc.	200	Ink jet	-	-	With alphanumerics
Gould Inc.	4822B	Electrostatic	11 wide	-	Raster scan
Gould Inc.	5105B	Electrostatic	22 wide	-	Raster scan
Hewlett-Packard	7221A	Pen	11 by 17	4,600	Flatbed, 4 colors
Houston Instrument	DP-11	Pen	11 wide	3,995	Drum
Houston Instrument	DP-8	Pen	36 wide	9,500	Drum, 3 colors
Phillips	PM-8125	Pen	10 by 15	2,195	XY Recorder, flatbed
Tektronix	4662	Pen	10 by 15	-	Flatbed
Versatec	1600A	Electrographic	-	9,950	Raster scan
Zeta Research	1200	Pen	11 by 14	10,000	Drum

For flatbed and drum plotters, the pens play an important role in determining the plot quality. Very fine lines are difficult to obtain with prolonged use of either felt-tip or ballpoint pens. Liquid ink pens can achieve a high quality, but are subject to clogging. Some models of plotters can accommodate several types of pens so that the user can determine line quality. Many manufacturers also have multiple pens that produce multicolor plots.

The quality of the plotted lines depends on the ability to draw smooth circles and arcs. For flatbed and drum plotters, the curves are made from a series of short straight-line segments. The quality of the curve increases as the segments become smaller, but each segment requires a vector command. The speed of the plot is decreased as the number of these commands increases. The ability to draw very short line segments works in contradiction to the ability for high-speed plots. Likewise, the quality is dependent on the system accuracy, resolution, and repeatability. The accuracy is the tolerance to which a point may be plotted. The resolution is the smallest distance that can be drawn between two points. Repeatability of the system is the ability of the plotter to return to a position that has already been plotted. Most high-quality pen plotters currently available have accuracies and repeatabilities ranging from 0.001 to 0.005 in.

10.2.4 Trends and Projected Developments in Printers and Plotters

With few exceptions, printers employ well-developed and time-proven technology. While the actual printer design may be different or new, chances are the technology is well established. The most evident exceptions to this are the laser xerographic and the ink-jet printers, both of which have been developed recently. The printer or plotter controller may incorporate a microprocessor, and while this method of implementation is relatively new, the microprocessor has been around long enough that the advantages of using it have been established. Table 10.2.4-1 presents the projected characteristics for printers in the 1980-1985 timeframe.

The low-speed impact printers, such as the cylinder, golf-ball, and daisy wheel, will see little or no improvement in speed, capability, or cost by 1985. All three have well-established markets and will continue to be used for low-speed mechanical terminals and office equipment. The number of these machines in use will increase slightly, although their popularity will decrease as more of the other printer methods come into wider use.

Drum, train, and band printers will all continue to dominate the computer printer market through the early 1980's. This is true due to the established momentum of the large impact printers and the lack of proven performance of the other alternatives. The number of large impact printers in use represents a major investment to industry and the replacement of these machines will be slow. The capabilities of the large impact printers will remain constant, with only minor increases in speed and reliability. This is due mainly to an increased vendor interest in the nonimpact printer methods since some of the latter are capable of higher speeds at equal cost and reliability. By 1985, the number of large impact printers will still be in the majority, but nonimpact technologies will be widely used and will hold a major portion of the market. These impact printers will remain the choice of users needing moderately large throughput (3,000 lpm) with multiple copy capability.

TABLE 10.2.4-1. PROJECTED DEVELOPMENTS FOR PRINTERS

TYPE OF PRINTER	SPEED	USE	COMMENTS
Cylinder Golf-Ball Daisy Wheel	Up to 70 cps	Office machines Time-shared computer terminals	Replaced by matrix and nonimpact printers
Drum Train Band	Up to 3,000 lpm	Medium-speed computer output	Some replacement at low speeds by matrix and electrographic printers Some replacement at higher speeds by electrostatic printers Will continue to be most popular medium-speed computer output
Matrix Impact	600 cps	Office machines	Will replace cylinder, golf ball, and daisy wheel
	1,000 lpm	Medium-speed computer output	Will replace slower drum, train, and band printers
Thermal-Matrix	300 cps	Office machines Status logging	Limited to slow speeds with low volume
Electrographic	5,000 cps	Office machines Status logging Small terminal hardcopy	Capable of medium speeds but limited in output volume by high paper cost

TABLE 10.2.4-1 - Concluded

TYPE OF PRINTER	SPEED	USE	COMMENTS
Electrostatic	3 to 5 sec/page	Terminal hardcopy	Widespread use for CRT terminal hardcopy capability
	500 lpm	Desk-top printers	Use for minicomputer printer
	5,000 lpm	Free-standing printers	Use for medium-speed computer output with printer located in work area
	20K lpm	High-speed computer output	Use for very-high-speed output from large computer center
Xerographic	30K lpm	High-speed computer output	Use for very-high-speed output from large computer center
Ink-Jet	50 cps	Office products	Use for office products and limited use for terminal output
	50K lpm	High-speed computer output	Use for very-high-speed output from large computer center

The use of matrix impact printers will increase for all applications requiring slow (300 cps) to medium (500 lpm) print speeds. The serial-matrix printers will gain widespread acceptance, with many applications in point-of-sale devices, equipment logging, and office products.

By 1985, the resolution of the small matrix printers will improve and the print quality will approach that found in the low-speed serial impact printers. Reliability will increase and be comparable with the serial-template impact printers by the early 1980's. The price of the serial impact printers will decrease slightly in this timeframe. The line-matrix impact printers will gain wide acceptance as medium-speed desk-top printers by the early 1980's. These machines will provide higher character resolution, reliability, and print speeds. Print speed should increase to 500 lpm by 1980 and to 1,000 lpm by 1985. The line-matrix printer will replace the larger line-template printers in many applications. These printers will be placed in the work environment since the noise level is much lower than for template impact printers. The line-matrix printer will find wider use in multicopy form printing and applications not requiring the higher speed of electrostatic printers.

For low speed (less than 100-cps) print applications not requiring multiple copy capability, thermal-matrix printers will find widespread use due to their quietness of operation and low cost. The higher cost of thermal sensitive paper will remain a disadvantage, however, and will restrict the thermal-matrix printer to applications with a limited amount of throughput. The speed of these printers will increase to the range of 300 cps, but the impact of the increase will be minimal since acceptance of the printer will still be limited by paper costs.

Electrographic printers will remain attractive for medium- to high-speed serial printing applications not requiring multiple copy capability. The very low cost, simplicity, and reliability of design will aid in widespread acceptance. Improvements in the paper will create better acceptance of the printed output. The rate of throughput will increase to 5,000 cps during the 1980-1985 timeframe, but the cost of the special paper will

restrict its widespread use as a high-rate computer output device except for hobby-home applications when the printer will be operated only a limited amount of time.

The direct electrostatic printer will see a major increase in use as a computer output device. The low-speed electrostatic printer will gain popularity as a hardcopy device for CRT terminals. In 1977, several companies introduced electrostatic printers for this application. Typical models take from 10 to 20 sec to form a page, but internal printer memory allows the information to be loaded in 1 to 2 sec, with the actual printing done "off line". The speed of these terminal printers will increase, with times of 3 to 5 sec per 8.5- by 11-in. page being common. General quietness of the CRT/electrostatic printer combination will make its use quite common in the office environment. The higher-speed electrostatic printers will provide the greatest competition for large impact line printers. Electrostatic printers are presently equal to impact line printers in terms of cost for a given speed. Electrostatic printers have the advantage of operating at higher speeds, however, and will be favored for operations requiring a throughput higher than can be obtained by impact printers. Additionally, the quietness of operation will allow the placement of printers in the work area. Desk-top electrostatic printers will be common in the early 1980's, with print speeds of 300 to 500 lpm. Free-standing office models will be capable of operation at 5,000 lpm, with the 132-character print line compressed to fit on an 8.5-in. format. Very-high-speed (10K to 20K lpm) electrostatic printers will be common as output devices for large computer centers. The number of electrostatic printers in use should not exceed the number of impact line printers during the 1980-1985 timeframe, mainly because of the number of impact printers already in use. The number of electrostatic printers will form a significant percentage of the printer market, however.

The newer-technology, very-high-speed printers will find increased use for high-speed computer output applications. The laser xerographic printers are currently being developed by several sources, and present

speeds range from 18K to 21K lpm. The reliability of these machines has been less than optimum, but the general reliability will increase in the early 1980's as the manufacturers have enough time and experience to improve the systems. The print quality will increase for the higher-speed machines, with the resolution becoming comparable with impact printers. The print speed will increase moderately by 1985, but not by more than a factor of 1.5. The high cost, high maintenance requirements, and the need for skilled operators will restrict the use of these machines to advanced computer centers where high speed throughput is essential.

Like the xerographic process, the ink-jet printer will gain increased use in both high- and low-speed applications. The low-speed ink-jet office printers will increase in speed and reliability as the technology matures. Some of the low-speed models already have excellent print quality, and it is unlikely that much improvement in print quality will be gained in the early 1980's. The main improvements will be in speed and cost. On the high-speed ink-jet systems, print quality will improve and the amount of required maintenance will decrease. Several new companies will enter the high-speed ink-jet market by 1985. The speed of these printers is not likely to increase during the early 1980's, however. The high speed of these machines along with the necessity for skilled operations personnel will limit their use to very-large-volume output applications at large computer centers.

The major trend in plotters will be toward use of microprocessors for internal controllers. Studies indicate (Ref. 10-4) that by 1980, over 80% of the plotting systems will operate either on-line or by remote time-sharing. This will be possible because the internal processor will relieve the control burden from the CPU and will reduce the amount of data that must be sent over remote lines. The plotters using pens for drawing will increase in multicolor capability, with three or four colors being common by the early 1980's on both large and small machines. Electrostatic plotters will increase in popularity because of their capacity for high-speed operation and high resolution.

Flatbed plotters will find use primarily as slower-speed off-line devices. Small electrostatic plotters will be used for terminal hardcopy devices. Medium-sized plotting applications will use drum plotters for multicolor graphs and electrostatic techniques for single-color and high-resolution graphs. Large output applications will use high-speed electrostatic plotters.

REFERENCES - SECTION 10

10-1 "Terminals: Smart and Dumb", Digital Design, pp. 60-65, December 1977

10-2 J. W. Schwartz, "Twenty-Five Years Without Panel TV", 1976 Society for Information Displays International Symposium Digest of Technical Papers, pp. 124-125, May 1976

10-3 "All About Computer Output Microfilm (COM)", Datapro Research Corporation, 1975

10-4 Sam Davis, "Hardcopy Computer Output - Recent Designs Feature Reliability, Intelligence, Lower Cost", EDN, pp. 38-54, October 20, 1977

BIBLIOGRAPHY - SECTION 10

E. W. Greeneich, "Thin-Film Video Scanner and Driver Circuit for Solid-State Flat-Panel Displays", IEEE Transactions on Electron Devices, pp. 898-902, July 1977

J. N. Orr, "Computer Graphics", Mini-Micro Systems, pp. 46-71, February 1978

"All About Alphanumeric Display Terminals", Datapro Research Corporation, 1976

"EL Displays Look Good to British", Electronics, pp. 36-40, July 7, 1977

K. O. Fugate, "Thin Film EL-Prognosis for Panel Displays", Technical Digest of 1977 IEEE International Electron Devices Meeting, pp. 69-73

Alan Sobel, "Summary: New Techniques in Video Displays", Proceedings of the Society for Information Displays, Vol 17/1, pp. 56-59, First Quarter 1976

"Displays", Electronic Engineering, pp. 69-79, February 1977

"SID '77 Will Show You Things You Never Saw Before", EDN, pp. 25-26, April 5, 1977

S. E. Scrupski, "Digital-Display Manufacturers Strive to Refine Their Technologies", Electronics, pp. 96-103, July 7, 1977

W. L. Buehner, et al., "Application of Ink Jet Technology to a Word Processing Output Printer", IBM Journal of Research and Development, pp. 2-9, January 1977

"Mini-Computer Controlled COM Recorder", Stromberg Datagraphix, Inc., San Diego, California

T. P. Brody, "A 6 × 6 Inch, 20 Lines/Inch Electroluminescent Display Panel", Conference Record of 1974 IEEE Conference on Display Devices and Systems, pp. 129-131

A. G. Dewey, "Projection Systems for Light Valves", IEEE Transactions on Electron Devices, pp. 918-929, July 1977

"A CRT Hardcopy Printer that is the 'Cat's Meow'", Mini-Micro Systems, p. 32, January 1978

H. O. Peprnik, "A Light-Reflecting Electromagnetic Display", 1974 Society for Information Displays International Symposium Digest of Technical Papers, pp. 136-137

W. E. Good, "Recent Advances in the Single-Gun Color Television Light-Valve Projector", Society of Photo-Optical Instrumentation Engineers - Simulators and Simulation, Design, Applications, and Techniques, Vol 59, pp. 96-99, 1975

J. McDermott, "Liquid Crystals, Lasers, and μ Ps Reflected in Tomorrow's Displays", Electronic Design 10, pp. 42-45, May 10, 1977

A. I. Lakatos and R. F. Bergen, "TV Projection Display Using an Amorphous-Se-Type Ruticon Light Valve", IEEE Transactions on Electron Devices, pp. 930-934, July 1977

D. B. Hakewessell, "7000 Lumen Color-TV Projector Development", 1974 Society for Information Displays International Symposium Digest of Technical Papers, pp. 138-139

"Hardcopy System", EDN, p. 120, October 20, 1977

L. T. Todd, Jr. and C. J. Starkey, IV, "High Brightness, High Resolution, Projection CCRT", Technical Digest of 1977 IEEE International Electron Devices Meeting, pp. 80A-80D

P. J. Klass, "Navy Tactical Situation Display Readied", Aviation Week & Space Technology, pp. 42-44, January 2, 1978

"All About Graphic Display Devices", Datapro Research Corporation, 1974

M. Bechtler and H. Krüger, "Dim Light is no Turnoff for Fluorescence-Activated LCD", Electronics, pp. 113-116, December 8, 1977

W. E. Howard, "Electroluminescent Displays", Conference Record of 1976 IEEE Conference on Display Devices and Systems, pp. 9-13

I. L. Wieselman, "Printers", Mini-Micro Systems, pp. 52-86, January 1978

A. L. Dalisa, B. Singer, and R. Liebert, "Matrix Addressed Electrophoretic Displays", Technical Digest of 1977 IEEE International Electron Devices Meeting, pp. 74-77

"Prototype LCD Image Generator Vies with Mechanical Flatbed Plotter", EDN, pp. 20-24, October 20, 1977

C. Suzuki, T. Uno, and S. Mito, "LED Color Display with Inherent Memory", 1974 Society for Information Displays International Symposium Digest of Technical Papers, pp. 80-81

"Japanese Showoff Liquid-Crystal Television Screen", Electronics, pp. 41-42, May 26, 1977

"Computer Plotting Equipment", Digital Design, pp. 48-50, December 1977

A. Sobel, "Approaches to Flat-Panel TV Displays", 1976 Society for Information Displays International Symposium Digest of Technical Papers, pp. 158-159, May 1976

"French Lab Produces Tiny LCD for Videophone Applications", Electronics, pp. 55-56, December 22, 1977

"Dense-Plasma Panel Users Must Seek New Sources", Electronic Engineering Times, p. 22, February 20, 1978

M. G. Craford, "Recent Developments in Light-Emitting-Diode Technology", IEEE Transactions on Electron Devices, pp. 935-943, July 1977

S. Mito, C. Suzuki, Y. Kanatani, and M. Ise, "TV Imaging System Using Electroluminescent Panels", 1974 Society for Information Displays International Symposium Digest of Technical Papers, pp. 86-87

W. M. Newman, "Trends in Graphic Display Design", IEEE Transactions on Computers, pp. 1321-1325, December 1976

A. Sobel, "Gas-Discharge Displays: The State of the Art", IEEE Transactions on Electron Devices, pp. 835-847, July 1977

"Terminals: CRT, Graphic Display, & Printing", Digital Design, pp. 55-77, January 1978

R. G. Kaufman, "Lamp Phosphors and Color Gamut in Positive-Column Gas-Discharge Cells for TV Displays", IEEE Transactions on Electron Devices, pp. 884-890, July 1977

"Digital Addressing Cuts the Neck from a CRT", Electronic Design 1, p. 36, January 4, 1978

A. L. Dalisa, "Electrophoretic Display Technology", IEEE Transactions on Electron Devices, pp. 827-834, July 1977

H. Simpson, "Printers '77", Digital Design, pp. 29-53, October 1977

"Kodak KOM-80 Microfilmer", Eastman Kodak Company, Rochester, New York

B. L. Frescura, H. Luechinger, and C. A. Bittmann," Large High-Density Monolithic XY-Addressable Arrays for Flat-Panel LED Displays", IEEE Transactions on Electron Devices, pp. 891-897, July 1977

"Comp 80 Specifications", Information International, Culver City, California

A. Sobel, "Current Display Research -- A Survey", Technical Digest of 1977 IEEE International Electron Devices Meeting, pp. 64-68

W. C. Holton, et al., "Design, Fabrication, and Performance of a Flat Tube Display", Technical Digest of 1977 IEEE International Electron Devices Meeting, pp. 78-80

11. SOFTWARE

It is natural to think of a data processing system as hardware. The computer, disks, displays, etc., are all visible, tangible electromechanical components. These devices, connected together, form a data processing system that will have known and predictable characteristics: the computer processes a given amount of data per second; the display has a specified resolution; the disk permits the storage of a given amount of characters. To change these characteristics, it is necessary to use different hardware components.

It is less natural to think of a data processing system as software. Software is neither visible nor tangible, yet software controls the way data are processed. A computer system is nothing more than a group of components designed to execute a sequence of commands. That sequence of commands is called software. The software turns a computer into a data processing device and specifies to the computer the sequence in which functions will be performed. Unlike the hardware, software is "soft"; the command sequence can be replaced easily with a new command sequence.

Software is normally organized into two main groups (Figure 11-1). The first group provides direct control of the machine and its peripheral equipment (disks, displays, card readers, etc.) and provides an interface between the specialized command sequences (called a program in this report) provided by a user and the computer itself. This group is called systems software and includes an operating system, language translators, and utility routines. Systems software is an extension of the hardware. Its purpose is to relieve the user of the necessity of understanding the detailed operation of the hardware.

Programs that make the computer perform specific, productive work for the user are called applications programs, and the collection of these programs is called applications software. Applications software is further classified as commercial or scientific software, depending on the area of application. Commercial software is used to automate business

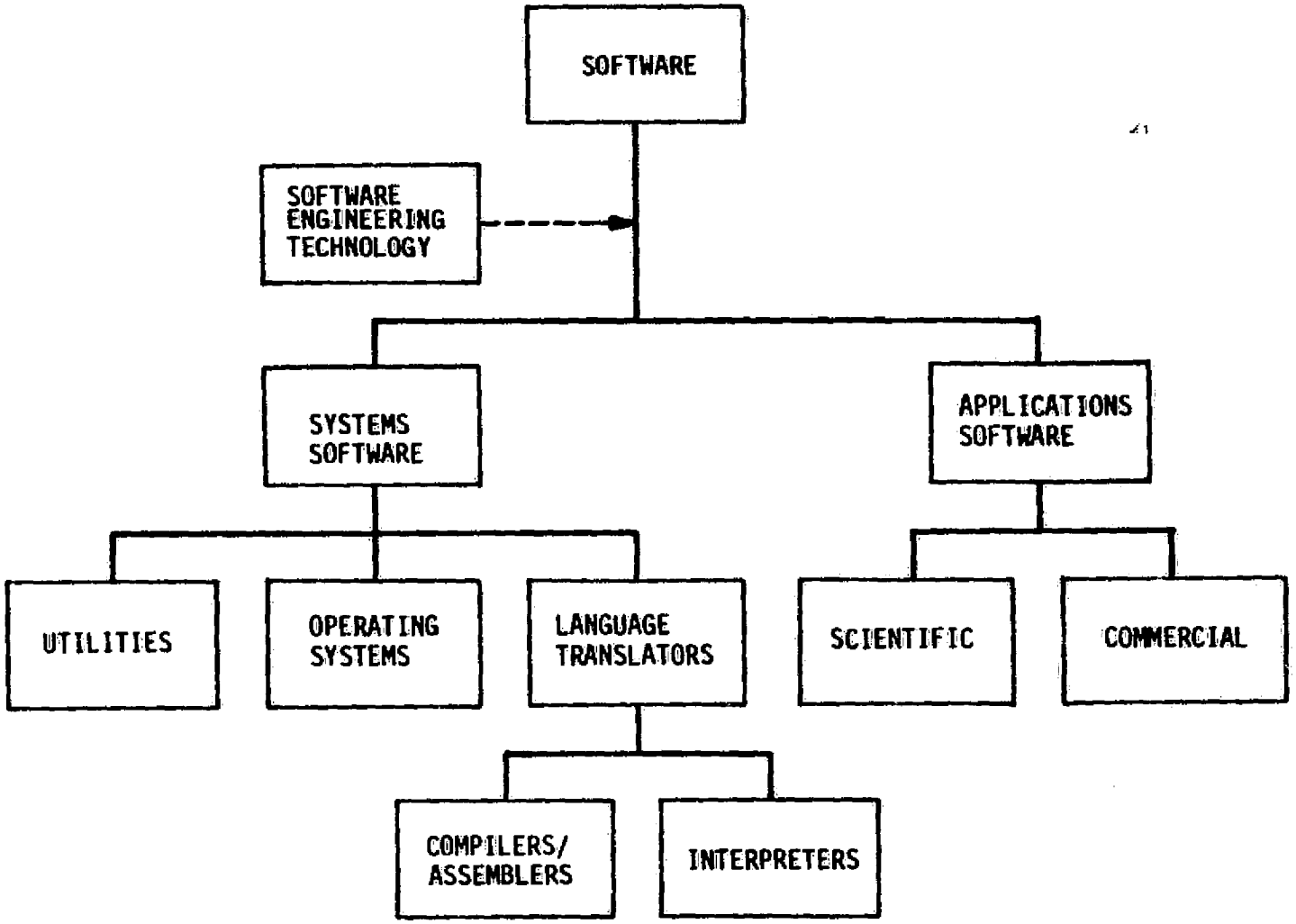


FIGURE 11-1. SOFTWARE ELEMENTS OF DATA PROCESSING SYSTEMS

processes and facilitate their operations. Scientific software is generally produced for scientific or engineering applications. These applications range from simple (a few commands) to highly complex (almost one million commands).

Software engineering technology (Figure 11-2) is centered around a software life cycle consisting of three phases: Initiation, Development, and Operation and Maintenance (Ref. 11-1).

The Initiation Phase begins with the recognition that a new computer application is required. The phase is performed primarily by the user or his representative. During this phase, the new application is studied to determine whether it is feasible and cost-effective to proceed with the development. Sometimes, depending on the size and complexity of the application or the user's skill, a number of additional studies are performed to support the feasibility of the application. These studies include detailed cost estimates, preliminary designs, scheduling, risk assessment, and options. Today, almost always, a document is produced defining the purpose of the application and its requirements at the end of this phase.

The Development Phase is the building phase. During this phase, the concept defined in the Initiation Phase is translated from an idea to a computer program (or a system of programs). The translation process may be subdivided into four steps:

- Requirement Specification - The concept defined in the Initiation Phase is analyzed and a set of software requirements is developed.
- Design - A solution that meets the software requirements is developed and further refined into the design of the program. The design is then checked against all the requirements to ensure that it satisfies them.
- Coding and Integration - The design is translated into a computer program. As each segment of code is written, it is tested first by itself and then integrated with other existing portions of the program and retested to validate that it performs as expected.
- Acceptance Testing - The completed program is formally tested to validate that it performs as envisioned during the Initiation Phase and is ready to become operational.

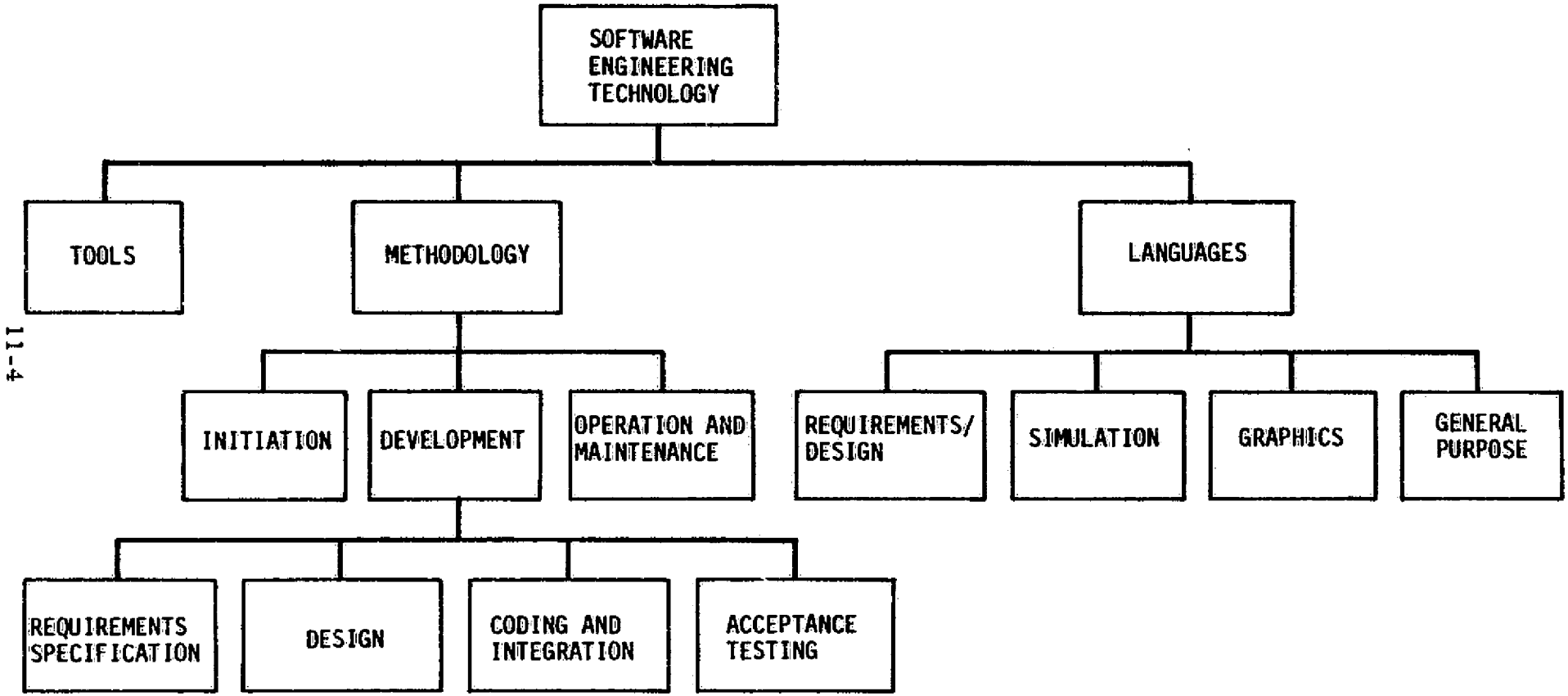


FIGURE 11-2. SOFTWARE ENGINEERING TECHNOLOGY

While the Initiation Phase is performed by the user, the Development Phase is generally performed by people who specialize in software. The overall process described in this phase is generally called programming and the people who perform the process are called programmers. Programmers are required since programming is a complex process and the user seldom has the skill or the time or personnel to develop the program. Often, due to the size of the program or short delivery times, or both, a group of programmers is required in the Development Phase. Generally, the user is relegated to the role of a monitor until he is asked to accept the program during Acceptance Testing.

After the program passes Acceptance Testing, it is turned over to the user and enters the Operation and Maintenance Phase of its life cycle. During this phase, the program is operational. The length of the Operation and Maintenance Phase varies from program to program and may last for a single execution or as long as it is still compatible with the user's computer. The Department of Defense, for example, estimates that large command and control programs have an average life of 10 years, while another study of commercial applications programs (Ref. 11-2) found the average life of a program to be 18 months (excluding one-time programs), with no program lasting more than 33 months.

Throughout the Operation and Maintenance Phase, programming activities continue. This makes this phase different from the maintenance phase of hardware, when only preventive maintenance and repairs to nonworking equipment are made. In fact, various estimates indicate that 32 (Ref. 11-3) to 60% (Ref. 11-4) of all programming activity is involved with maintenance programming. In contrast, computer hardware maintenance costs 8 to 12% of the price of the machine.

This section of the report analyzes three aspects of software and and software engineering technology -- system software, programming languages, and methodology -- and attempts to define the state of the art, trends, and forecasts for 1985.

Hardware forecasting is comparatively a more objective process than software technology forecasting. This is because a hardware forecast is based on establishing trends using specific and well-defined metrics, such as cost per bit, add time, transfer rate, byte capacity, and size. However, there are no comparable meaningful metrics in software yet. In fact, there is little agreement on even the definition of the metrics. As a result, historical software data are either lacking or impossible to interpret. Therefore, software forecasting is a subjective process, based on information extracted from scientific papers, books, usage surveys, analogies to related fields, etc..

One additional principle can be applied to software forecasting. Known as the "7-year rule", it states that a minimum of 5 to 10 years is needed to bring technological innovation from initial laboratory demonstration to general availability in practical applications (Ref. 11-5).

11.1 OPERATING SYSTEMS

The term "operating system" denotes an organized collection of programs (implemented in either software, firmware, or hardware) and data which have as their purpose the management of a computer system's resources and the control of programs executing on that system. This section consists of four subsections addressing, respectively, the development, current state of the art, current trends, and future development of operating systems. A concise historical summary is presented first to provide the context for the following subsections.

11.1.1 Historical Summary

The earliest users of computer systems not only programmed their application but also operated the computers. They set toggles, fed cards, and watched over the system until the run was completed. This grossly inefficient use of the programmer's talent was corrected by employing a person trained in hardware handling (e.g., control panel wiring, forms loading, tape mounting) to operate the computer. The user (programmer) provided the operator with his program and with written instructions telling the operator exactly what actions were needed to process his work. The idea of automating part of the operator's job by allowing automatic interpretation of some of the operator's instructions led to the development of operating systems.

The first-generation operating systems (often called monitors) were simple control programs providing job-to-job transition by sensing end-of-job messages and branching to the next job from the card reader. Each user's job had control of all the system's resources during the job's execution. Card reading/punching and printing were performed on line.

As computers became faster, the second-generation operating systems evolved. The speed disparity between the CPU and the I/O devices led to the development of batch processing. This concept increased the performance of the main computer by exchanging low-speed I/O devices for higher-speed devices. The slow card-reading, punching, and printing operations

were moved off line. The simple monitor system was further improved through the concept of the I/O channel and the CPU interrupt capability. The I/O channel allows I/O and computing to proceed concurrently, and the CPU interrupt allows an external event to interrupt the execution of the CPU. Implementation of these capabilities resulted in complex operating systems, part of which were required to be resident in main storage at all times. The main storage resident portion of the operating system provided intercommunication and control among operating system routines and user programs.

The effectiveness of certain I/O devices was improved by using blocking to combine several logical records into one physical record. The functions of blocking and deblocking were provided by the operating system. The technique of multiple buffering was also used. The second-generation operating system fully supported sequential, indexed-sequential, and direct file organization. A significant advance was made when operating systems began to allow symbolic device assignment. This was the advent of "device management", one of the primary functions of an operating system.

Chaining was an early technique that provided the capability for running programs requiring more than the available main storage. This concept was replaced by the more easily used concepts of overlays and labeled commons.

Even with these capabilities, the second-generation operating system was essentially a collection of independent programs that were accessed one after the other to perform the functions for which they were designed. The relatively low degree of integration was primarily a result of the limited interrupt capability of the hardware, limited memory protection, and lack of I/O channels. The typical operating system ran in a serial batch mode, one program at a time.

The computer's third generation was ushered in with the introduction of integrated circuit logic. With these, the speed disparity between the computer and its I/O devices became even more pronounced.

The effort to obtain maximum utilization of the hardware led to multiprogramming. Under this concept, the CPU is shared by a number of main storage resident programs. An early basic fixed-partition multiprogramming operating system used four partitions containing, respectively, the supervisor routine, a primary foreground function (e.g., terminal support), a secondary foreground function (e.g., I/O handlers), and a background function (e.g., batch job stream processing). An advanced fixed-partition system included as many as 16 partitions of varying sizes. The system used privileged instructions to protect each partition's storage area.

These operating systems were lacking in at least two areas. First, they normally did not provide roll-in/roll-out capability. That is, once a program became resident in a partition it ran to completion (or termination by the operator). Second, they lacked a resource manager. Thus the computer operator was responsible for doing job planning, computer utilization planning, and balanced production scheduling.

The next step in the evolution of operating systems was the development of a system that overcame problems inherent in fixed-partition systems by dynamically allocating main storage. Resource management was implemented, in part, through the use of priority schemes. Priorities were used to determine not only which job was processed next but which job should be given control of the CPU next. The ability to suspend a job, roll it out to secondary storage, and later roll it back into main storage for continued execution was available. This allowed low-priority jobs to be suspended so that higher-priority jobs could be hurried through the system.

Computers became even faster as electronic technology advanced, causing the speed disparity between the CPU and its I/O devices to be further aggravated. The main storage capacity available limited the number of active programs that multiprogramming systems could handle. Once again, CPU utilization became unacceptably low. It was possible to increase the number of active programs without increasing the amount of main storage by breaking each program into segments and using the old roll-in/roll-out logic to allow only selected segments of the program to

be resident in main storage. The segments not currently being used were stored in secondary storage. Since the roll-in/roll-out operations (called paging) provided by the operating system are transparent to the user, this "extension" of main storage is called virtual storage. Although this main storage management technique is not a recent innovation, it is still considered a state-of-the-art technique.

Another type of operating system is the time-sharing operating system. This type of operating system has many users, each treating the computer as if he were the exclusive user. In this environment, the users usually interact with the system through remote terminal devices. Using the technique called time-slicing, the operating system passes control of the CPU to each user in turn, but limits each user to a short period, a time slice, of control. To balance the system load and provide all users with adequate response time, the amount of time in the time slice given to a user may be made a dynamic function of several parameters, such as overall system load and past resource demands of the user.

This discussion has presented the major innovations that have led to operating systems as they exist today. A modern operating system will integrate several of these and other techniques into a coherent system to effectively manage the system resources. For example, it may provide batch processing by using dynamic partitioning, multiprogramming, and virtual memory, while providing time-sharing services using a time-slicing technique. It will manage several different types of devices (tape drives, disks, drums, telecommunication) using different management algorithms and buffering techniques for each. Many system control parameters will be of a dynamic nature so that the operating system may continually make adjustments in its use of resources. Further, more detailed capabilities of some currently available operating systems are presented in the following subsections.

11.1.2 State of the Art in Operating Systems

To give the reader a broad view of the capabilities of currently available operating systems, summaries of the capabilities of several modern operating systems are presented.

11.1.2.1 IBM VM/370 - The IBM VM/370 can be classified as a fourth-generation operating system. It permits one machine to appear to be many at any time. The illusory effect is the result of using the virtual machine concept. A virtual machine is a simulated functional equivalent of a real system/370 that consists of a virtual CPU, virtual storage, virtual I/O devices, and a virtual console.

To employ VM/370 facilities, each user defines his unique machine configuration and requirements plus the operating systems to be used, which can be either the OS/370 or OS/360. This description is recorded in a directory held on a VM/370 file. When a user calls for his "system", the definition is retrieved and VM/370 matches his virtual machine I/O device addresses with real device addresses. Real resources are allocated to users in a computer-sharing mode, and each user is given frequent CPU time slices until he is finished.

VM/370 is both an operating system and a multi-access time-shared facility that permits users to develop virtual machines. It consists of two major components: the control program (CP) and the conversational monitor system (CMS). Through these, the user creates and controls the virtual machine he desires. The CP manages the real resources of the system, including CPU time, to create and control multiple concurrent virtual machines that can run under different operating systems. CMS operates under the direction of the CP and provides a general-purpose conversational time-sharing capability suitable for problem solving, program development, and general conversational work.

Work done by each virtual machine is scheduled and controlled by the selected OS running under VM/370. The time and system resources required to do the work are managed by the virtual machine facility. The following advantages are claimed by IBM for the virtual machine concept:

- Programming cost reduced
- Memory fragmentation eliminated
- Enhanced on-line application possibility
- System backup assured.

11.1.2.2 CDC SCOPE and KRONOS - CDC has developed two operating systems for its medium to large-scale machines. They are SCOPE and KRONOS. SCOPE is designed for generalized multimode processing environments; it supports batch, remote batch, time-sharing (up to 80 terminals), real-time, and interactive graphics. KRONOS is designed for large communities of asynchronous terminals in a time-sharing environment; it supports time-sharing (up to 512 terminals), batch, remote batch, and deferred batch.

Although it supports multimode operation, the primary purpose of SCOPE 3.4 is to support batch and remote batch. All other modes compete with remote batch for the resources. However, remote batch is given a higher priority and a longer time slice for both memory residency and CPU execution.

SCOPE 3.4 handles time-sharing in the following manner. The Peripheral Processor Unit (PPU) takes care of the utility functions such as log in, status, open files, etc. Utility functions are completed before the PPU passes a job to the disk input queue. In this way, a PPU can be viewed as the front-end processor for the time-sharing jobs.

KRONOS was designed to efficiently handle time-sharing and batch processing. Deferred batch is a new capability in KRONOS, but remote batch capability is limited.

The scheduling of a job in both operating systems is determined by a four-digit priority classification. The first two digits are declared by the user. The last two digits are determined by the system scheduling algorithm. The resulting priority decides the memory residence and CPU time slice. The scheduling algorithm, in general, favors I/O bound jobs over computationally bound jobs.

To accommodate the changing loads with time on different parts of a multimode environment, the KRONOS operating system provides a tuning

capability. The system can be tuned by setting parameters during system generation time. The parameters can assign different priorities, memory allocations, time slices, etc., which in turn affect the entire system response and performance.

A new operating system, Network Operating System (NOS), will eventually replace both SCOPE and KRONOS. The release of NOS will be through several stages. NOS is designed for batch, remote batch, deferred batch, time-sharing, transaction processing, real-time, and multiple mainframe system functions between the PPU-based monitor and the CPU. Unlike SCOPE or KRONOS, the CPU will have the responsibility for memory management and direct-access device data management.

NOS will support up to 25 dynamic partitions. One partition is used for each of the application executives and one is used for I/O spooling. It provides the computer with a networking capability by the addition of a network supervisor which is another module under NOS. It will also operate as a working processing resource in a computer network and keep track of the processing resources available on the network as well as communication paths available to that resource.

11.1.2.3 Burroughs MCP - Burroughs Corporation produced the MCP V (Master Control Program) for its medium to large-scale computers. The MCP V is a multimode operating system supporting batch, time-sharing, and transaction processing. It requires 30 to 32 Kbytes of resident memory. The normal mode of operation of MCP V is multiprocessing. It includes the following functions: initiation of programs on a priority basis, dynamic storage allocation on both main memory and disks, virtual memory operation using a memory segmentation technique, and multiprogramming.

The structure of MCP V is a skeleton program with a number of overlays, but only one copy of a particular overlay can be in memory at any time.

An interesting feature in MCP V is that it includes a new scheduling algorithm that attempts to solve the most critical scheduling problems (on any computer). It uses a three-level system of priorities; system schedules

are based on a combination of memory priority, process priority, and schedule priority. Memory priority refers to the decision by the system as to which program currently in memory should be suspended to make room for higher-priority programs. Process priority determines the best way to enhance system throughput to dynamically adjust its loading and processing. Schedule priority is a user input. The three-level system of priority does not guarantee the best solution to scheduling problems, but it is an interesting attempt.

MCP V also permits memory sharing, which is a technique whereby multiple copies of the same program share a common area of memory. It also supports shared-disk operations.

A new version, named MCP VI, has a completely different architecture and provides a host of new functions. MCP VI is structured like a series of independent programs so that more than one copy of the same program can be in memory functioning concurrently. Although MCP VI is more dynamic than MCP V, upward compatibility is maintained between the two operating systems.

In MCP VI, the relationship between a job and a task is very much like that in a PL/I program. A job/task manager groups tasks into jobs and allows different types of task handling (e.g., a job may run some synchronous and some asynchronous tasks).

In a multiprocessing environment, each processor can be under the control of different versions of the operating system. A single copy of "parent" MCP VI can reside on disk, and its overlays can be shared by various processors. Alternatively, independent versions can reside on private, processor-associated disks.

Overall, the major design goals of the MCP operating systems are to optimize the hardware capabilities and increase the job processing rate of the system.

11.1.2.4 Xerox CP-V - CP-V operates in a multimode environment. It supports batch, remote batch, deferred batch, time-sharing, transaction

processing, and real-time. It was designed for Sigma 9 multiprocessor systems. One CPU is designated to handle I/O operations and schedules, execute user programs, and service all user requests. All other CPUs are secondary and serve as computing peripherals. Multiprocessing is transparent to users, with all processors contributing time to the execution of a user job.

CP-V is a single integrated operating system with a single monitor program to schedule and operate all system modes. It is a hybrid and extension of XOS and UTS (Universal Time-Sharing), with emphasis on UTS.

One of the unique features in CP-V is its dynamic tuning capability. There are up to 80 parameters that can be changed from the configuration console at any time by the operator. These changes can affect time slice and size of memory partition. In addition, privileged users can be given access to change partition size to permit oversized programs to run.

Up to 16 batch partitions are permitted, which are set to service batch jobs with different main memory requirements. The size of the jobs serviced by each partition can be different and can have unique time slices for execution. The intent of the different main memory resource and time-slice assignments is to ensure that all programs will be given a chance to execute regardless of memory requirements.

Two advanced features in CP-V are the following:

- Users can address most of the system facilities from any mode. This is because all modes use the same data.
- The transaction processing (TP) system allows users at remote terminals to perform business processing transactions. TP handles communication lines, preliminary edits, formatting, and message queuing and calls the appropriate application programs.

11.1.2.5 Bell Laboratories UNIX - UNIX is one of the operating systems designed for small-scale computers. Bell Laboratories developed UNIX to operate PDP-11/40 or PDP-11/45 computers. It is a general-purpose, multiuser, interactive operating system.

UNIX offers a number of features seldom found even in large operating systems. These features include:

- A hierarchical file system incorporating demountable volumes
- Compatible file, device, and inter-process I/O
- The ability to initiate asynchronous processes
- System command language selectable on a per-user basis
- Over 100 subsystems including a dozen languages.

The most important achievement of UNIX is that it demonstrates that a powerful operating system for interactive use need not be expensive either in equipment or in human effort. It is characterized by its simplicity, elegance, and ease of use. One thing worth mentioning is that UNIX is written in a high-level language C. This may be one of the reasons that contributes to its development cost saving.

It requires 42 Kbytes of main memory. The system calls to do I/O are designed to eliminate the differences between the various devices and styles of access. There is no distinction between random and sequential I/O, nor is any logical record size imposed by the system. The size of a file is determined by the highest byte written on it.

Communication with UNIX is carried on with the aid of a program called the Shell. It is a command line interpreter; it reads lines typed by the user and interprets them as requests to execute other programs. It also provides features of multi-tasking by allowing Shell, which itself is a recursive program, to respond to multicommands. The Shell is designed to allow the user full access to the facilities of the system since it will invoke the execution of any program with appropriate protection mode.

It is expected that the success of UNIX will affect many future operating systems for large-scale computers. The architecture of UNIX will lead the way to improve the coherence, simplicity, and elegance of operating systems.

11.1.3 Trends in Operating Systems

This subsection addresses trends in particular areas of operating system theory and practice. The following paragraphs discuss more general topics.

An outstanding pattern which is expected to continue is the operating system family trend. Several popular operating systems have been allowed to evolve through several generations with many releases in each generation. The systems have been improved, new features added, and even major changes made in the hardware while keeping the basic system intact and while holding the user's interface relatively constant. This trend has been and will continue to be encouraged by the high cost of both developing a new operating system and converting old software to run under a new system.

Another trend is the increasing use of high-level languages for implementing operating systems. For example, the operating system HYDRA has been implemented in the language BLISS and UNIX in the language C. Since the use of a high-level language decreases development costs and improves the maintainability of the system, this trend will continue (of course, operating systems for small microcomputers will probably continue to be written in assembly language).

11.1.3.1 Management of Main Storage - Computer systems having over 8 Mbytes of main storage are not uncommon today. Commercially available machines are available which support 16 Mbytes of main storage, and in the future, systems with over 100 Mbytes of main storage may become commonplace. It seems likely that as main storage expands, applications demand for storage will expand, as it has in the past, to demand even larger memories. Thus the management of main storage will continue to be important. In the past, as main storage expanded, the operating system also expanded, allowing more complex (and hopefully more effective) algorithms for main storage management (and other functions) to be implemented. This trend will continue. The relaxation of many constraints

of the past and better understanding of the necessary hardware structures will result in more effective and more efficient main storage management techniques.

11.1.3.2 Management of I/O Devices - The management of I/O devices, such as card readers, printers, tape drivers, disks, drums, plotters, etc., and their supporting devices, such as control units and control channels, will continue to be important. Techniques for the efficient use of I/O devices have included off-line peripheral operations, direct-coupled systems, attached support processors, and virtual systems. All these techniques may be thought of as methods for implementing virtual devices. The virtual device concept allows the execution of a program to be divorced from dealing directly with slow-speed I/O devices (e.g., the spooling technique can be used to make one physical card reader appear to be many virtual card readers). The present trend is toward even more virtualizing of devices.

The task of handling the physical devices can be more efficiently done if certain knowledge of the state of the device is available. For example, the IBM 370 provides rotational position sensing and block multiplexing, which were not available on the earlier IBM 360. There is a trend toward making more hardware facilities available to the software for better utilization of I/O devices.

11.1.3.3 Management of Information - The operating system modules that perform information management are sometimes collectively referred to as the file system. The file system is concerned only with the simple, logical organization of information into files or data sets. There are many similarities between the file system and the segmentation and paging concepts used in memory management. In fact, in MULTICS and other systems, the file system and the segmentation and paging mechanisms are integrated. The trend seems to be toward even more merging of modules of memory management and information management.

Recently, very-large-capacity on-line auxiliary storage devices called mass storage facilities have appeared (e.g., the SDC TBM II and

the IBM 3850 - see Section 8.3). These devices provide the capability to automatically access on the order of 400 billion bytes of information and to access it more quickly than was possible with previous forms of automatic mass storage. This continues the trend of increasing the amount of on-line information that must be managed. Auxiliary storage management is moving toward more automatic management of the different storage facilities, with the storage devices becoming more transparent to the user. To provide efficient access and economical long-term storage, data will be moved automatically, in accordance with storage management policies, from one device type to another. Backup of data, protection of data integrity, and data security will be provided by the operating system.

11.1.3.4 Proof of Correctness - Program proofs offer, in theory, a powerful way to obtain very reliable operating system software. Although there are no commercially available verifiers that can be used to test software for correctness, much of the motivation for the changes currently taking place in software design is based on the concept of correctness of programs (including both application programs and operating systems). Verifiers do exist in the universities, but they have not been verified themselves. It is also important to note that even if an algorithm is verified to be correct, this does not guarantee the correctness of a program written to implement that algorithm on any particular machine.

There are a number of powerful methods for proving that programs meet stated specifications. Many different types of programs, high-level language, assembly language, and parallel, have been verified. Progress is being made toward handling (verifying) more complex programs (Refs. 11-6 and 11-7), and techniques have been suggested which, in theory, allow proof of correctness of complex systems (Ref. 11-8). At present, a program is much more likely to be amenable to verification if it was constructed with verification as an original design goal. However, some experts in the field believe that the "right kind" of arbitrary program (a well-defined program, written in a suitable high-level language, etc.) can be verified if sufficient manpower and computer time are available.

11.1.3.5 Kernel-Based Operating Systems - The Kernel approach to the design of operating systems supports modularity and the separation of policies and mechanisms. The hope, of course, is that the resulting operating system can be proved and certified to be correct. Several operating systems, including HYDRA, Texas Instruments' DSOS, and the University of Rochester's ALEPH, have been implemented using the Kernel concept. The growing concern for providing computer security and protection in operating systems has also contributed to the emergence of Kernel-based systems. The most advanced protection features currently available seem to be present in the HYDRA Kernel (Ref. 11-9).

Problems such as Mutual Suspicion, Confinement, Conservation, and Limiting Propagation of Capabilities are successfully addressed by the HYDRA system. It also provides control protection and object-level access. In Reference 11-10, Linden surveys structures to support security and reliable software and concludes that there exists the potential for a breakthrough in both security and reliable software. The Kernel approach seems to be leading the way to this breakthrough.

11.1.3.6 Computer Networks - Several projects have been undertaken to study, design, and build network operating systems. A network operating system includes the software and protocols needed to allow a set of interconnected autonomous computers to be conveniently and effectively used together. Examples of network operating systems include the Resource Sharing Executive (RSEXEC) Project, supported by the Defense Advanced Research Projects Agency (DARPA), and the National Software Works (NSW) Project, supported by DARPA and the Rome Air Development Center. Although systems have been designed and built, there is a lack of attempts to measure or analyze the relative cost-effectiveness of alternative methods of implementing such systems.

At this time, network operating systems have not developed sufficiently for trends in details to be appearing. However, the general movement is toward having the operating system handle communications between elements of the network without the user knowing where the actual computer power or data reside.

11.1.4 Projected Developments in Operating Systems

The general outlook for operating systems in 1985 is for them to be using many of the same basic techniques (multiprogramming, time-slicing, buffering, etc.) that are used today. These operating systems will be implemented using a combination of hardware, firmware, and software and will be members of operating system families, with the software portion written (for the most part) in some high-level language. Current techniques will be refined to provide flexibility of application and dynamic response to load changes. The user will find the operating system more easily used and more dependable. Although these future operating systems will evolve from present operating systems, they will differ in that the operating system will be distributed among the system elements to a much greater extent than is currently being done. Because of the low costs projected for future processor/memory technology (see Sections 1 and 7), future large-scale computer systems will likely be implemented as some form of multiple processor with significantly greater throughput capacity than currently available. Future operating systems will therefore assume a greater role as managers of communications between the various system elements. The following subsections outline projected developments that will support these operating systems' characteristics.

11.1.4.1 Proof of Correctness - Although automatic and manually assisted program verification technology will probably not be capable of dealing with specifications and programs of the size and number involved in a complete commercial operating system, by 1985, formal verification of some major subsystems will be accomplished. Based on current developments (Ref. 11-11), it is not unrealistic to expect that, by 1985, there will be commercially available operating systems that include certified and verified security kernels. For verification to be accomplished, operating systems will be tightly designed (e.g., designed with verification as an objective), their functions precisely defined, and their implementation will be modular in nature.

11.1.4.2 Operating System Firmware - It seems apparent that more and more operating system functions will be performed by hardware/firmware. Factors pushing development in this direction include the continued decline in hardware/firmware costs, the increased flexibility of firmware (e.g., microprogrammable functions may currently be loaded from an external device), and the significant performance gains realizable by this procedure. Aided by architectural and software advances, manufacturers will be able to apply microprogramming systematically to the implementation of computer systems, including operating systems. Implementation of the most frequently executed primitives of an operating system nucleus in firmware is already being experimented with (Ref. 11-12). By 1985, at least these primitives will reside in (and be executed by) firmware, and perhaps more extensive portions of the operating system will be so implemented.

11.1.4.3 Computer Networks - Networks and network operating systems of some power will be common by 1985, but these will not yet have evolved into true distributed data processing systems. The problems, such as security, data integrity, and resource management, found in a central computer are magnified in a distributed processing environment. Of course, data transmission between sites and multiple file access become problems. These and other problems will be the subject of intense research, but they will delay the implementation of true distributed data processing systems well past 1985.

11.2 LANGUAGES

11.2.1 Introduction

This section evaluates computer languages, past, present, and future, in terms of their effect on programmer productivity. Computer languages are those tools that are used for expressing ideas as well as communicating with computers. For the purpose of developing the subject, the section is divided into the following general areas: overall history (1945-1978) and state-of-the-art and trends and projected developments in general-purpose languages, simulation languages, graphic languages, and the emerging field of system requirements specification/analysis languages (SRSAL).

This section does not attempt to provide a comprehensive study of all languages; it presents those languages whose primary objective is to be used as tools in the development of computer-based systems. Therefore, only a few languages in the general application-oriented language category are discussed.

The definitions given in Table 11.2.1-1 are used to conveniently discuss the subject without the need for references to each individual language. The four language classifications are: general-purpose, simulation, graphic, and system requirements specification/analysis. For discussing the last classification, the definition of "languages" has been broadened to include languages that do not currently produce executable code.

The most promising language areas (relative to increasing the productivity of programmers) are graphic languages, as described in Reference 11-13, and the development of system requirements specification/analysis languages. The reason for this is that these language areas in combination will facilitate both the understanding and analysis of computer system problems. The emergence of these two tools has been assisted by the increasing numbers of interactive graphic terminals.

TABLE 11.2.1-1. CLASSIFICATION OF LANGUAGES

LANGUAGES	DEFINITION
1. General Purpose Languages	Languages whose end objective is to produce object code (e.g., assemblers and compilers).
2. Simulation Languages	Languages used to model complex systems for analysis and detection of definition and/or design problems. Characterized by the fact that the end objective is not the simulation output object code.
3. Graphic Languages	Languages whose main objective is a higher level of man-machine (man-computer) interaction achieved by the use of graphic image detecting and/or producing devices. These languages are characterized by the fact that their primitives include two-dimensional graphic symbols in addition to the limited set of standard alphabetic, numeric, and special characters used in non-graphic oriented languages.
4. System Requirements Specification/ Analysis Languages	Languages whose primary objective is to describe the requirements for a total system or its parts in precise, unambiguous, and quantitative terms for detailed analysis. Included in this definition are those languages which claim the preceding characteristics.

11.2.2 History and State of the Art in Languages as They Impact Programming Environment, Techniques, and Productivity

Table 11.2.2-1 lists, in chronological order by year, the entrance of various language types, corresponding programming environments, and development techniques.

11.2.2.1 Statement of the Problem - Despite the proliferation of programming languages (over 600 were identified in Reference 11-14) and software system development techniques, these software tools seem to have no noticeable effect on increasing the productivity of programmers. This situation is reflected in the facts and figures in the following paragraphs.

According to one source (Ref. 11-15), "The cost of software in present and future data processing systems is becoming dominant, which is more than 50 percent and is expected to increase rapidly to a much higher percentage in less than a decade. (In some projections, e.g., [Reference 11-16], this figure reaches 90 percent.)" The reason for this appears to be that programmer productivity is not keeping pace with the developments in computer hardware.

In the following analysis "productivity growth factor" for the period year 1-year 2 is defined to be the ratio of source instructions produced in year 2 to instructions produced in year 1 with equal quantities of programmer effort. According to Reference 11-17:

- The productivity growth factor for 1955-1965 was about 2.1.
- The productivity growth factor for 1965-1975 was about 1.1.
- The greatest increase in productivity was during the period from 1955 to 1960, which reflects the effect of compiler-level languages on the computer industry at that time.

The productivity growth factor is an indication of the effectiveness of our software development tools on the production of computer

C-8

TABLE 11.2.2-1. HISTORY OF DEVELOPMENT TOOLS

YEAR	LANGUAGE TYPE	PROGRAMMING ENVIRONMENT	DEVELOPMENT TECHNIQUES
1945	Machine	Single-job	
1952	Symbolic Assembler	Real-time	
1953	Macro-Assembler		
1956	Compiler Interpreter		
1959	Operating Systems Job Control Simulation	Batch	
1961	String Processing		
1963	Graphic	Timing-sharing	
1965	List Processing		Configuration Management, Process Construction
1969	Requirements (Ref. 11-39)		Structured Programming, Modular Design
1970	Data Base Manipulation		Virtual Memory Top-Down Design Top-Down Implementation Top-Down Testing Chief Programmer Teams Step-wise Refinement Data Base Design Decision Tables
1972			Software Physics Verification and Validation Correctness Proof Software Reliability Symbolic Execution Structured Walkthrough Software Quality Control
1974			Software Engineering Technology
1975	Process Design		

systems. Therefore, it appears that the tools developed since 1960 are not solving the basic problems associated with software.

11.2.2.2 Reasons for the Problem - The stimulus associated with the above-stated problem has been the need for more functionally complex computer systems and the entrance of a batch-oriented programming environment. The reason that languages and techniques developed from 1960 to the present have not been effective in increasing programmer productivity is that these tools have been attacking the wrong problems. The problems that must be solved are:

- Lack of definitive, auditable steps in the programming process; management has no visibility of the programming progress
- Lack of discipline within the programmer community
- Lack of responsibility within the systems analyst community
- No major improvement in languages
- These languages provide no means for comprehensive analysis of the system definition and design ideas before full implementation of the system.
- These languages provide no means for communicating quantitative system requirements between systems analysts and programmers.
- These languages provide no means for communicating quantitative test criteria between systems analysts and independent test teams.
- A false sense of security has been created by each of the various technical fads developed during the latter 1960's and 1970's.

The industry has been attacking the wrong problem by looking for solutions (graphically illustrated in Figure 11.2.2.2-1) in "high-level languages", which do not provide much more advantage than those compilers produced in the 1955-1960 period. The following facts support the argument that the problem will be solved in the area of system

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

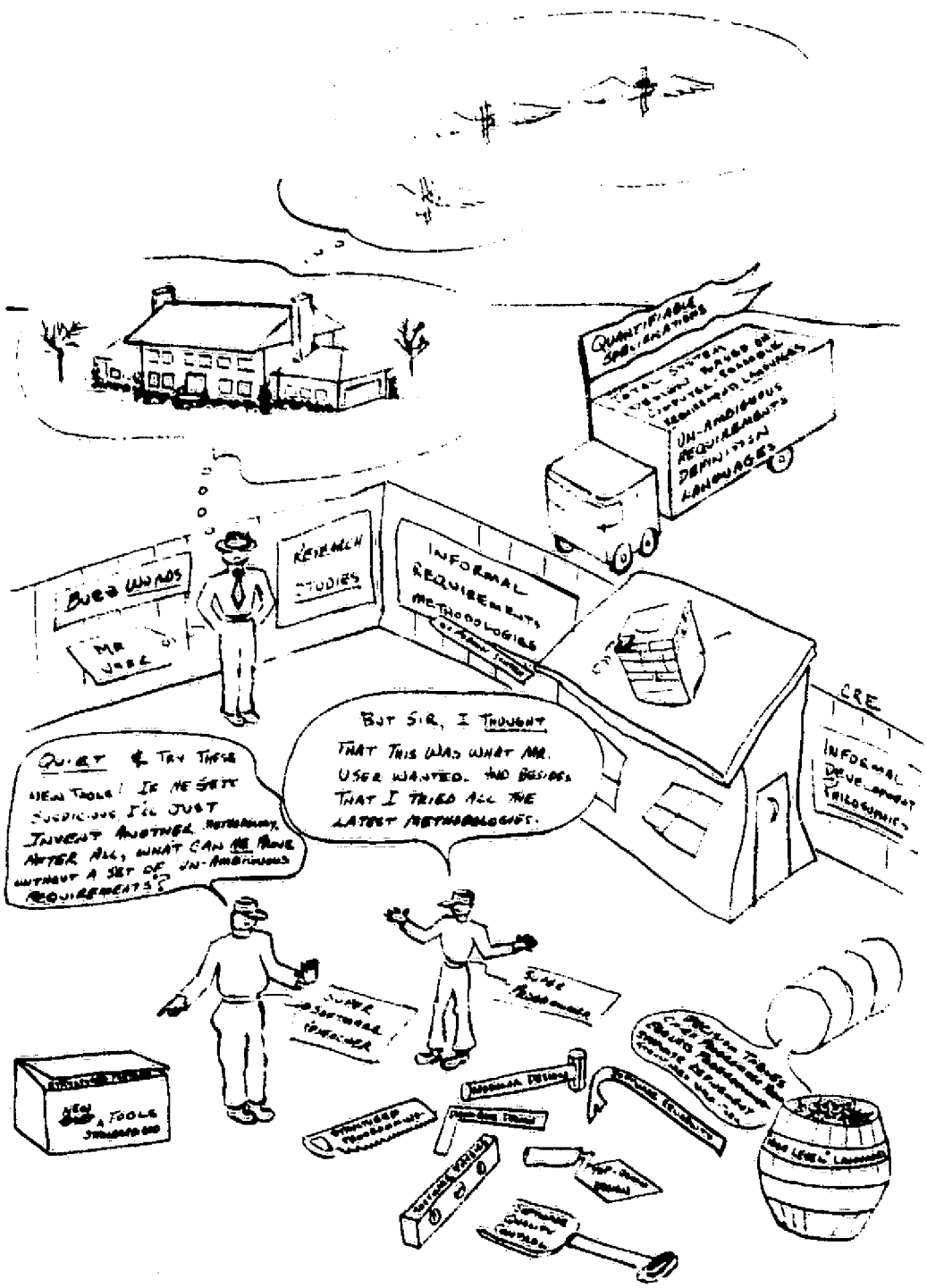


FIGURE 11.2.2.2-1. EFFICIENT METHODOLOGIES FOR FINDING THE WRONG SOLUTION

requirements specification and not in languages and techniques, which only support the limited area of coding:

- 36 to 74% of computer system problems, by count, are design errors, not coding errors (Ref. 11-18).
- Design errors require 1.5 to 3 times the effort to correct relative to implementation errors (Ref. 11-19).
- Therefore, 46 to 90% of the effort to correct all errors is associated with design errors, determined as follows:
  - Weighted Low (%) = $36(1.5)/[36(1.5)+64(1.0)]$ = 0.45763 (46%)
  - Weighted High (%) = $74(3.0)/[74(3.0)+26(1.0)]$ = 0.89516 (90%).

The following SAFEGUARD system example (7.5-year effort, Ref. 11-15) provides a specific example supporting the thesis:

- 20% of development time was spent on system engineering.
- 20% of development time was spent on design.
- 17% of development time was spent on code and unit testing.
- 43% of development time was spent on integration.
- Therefore, 83% of development time was spent on noncoding activities.

The fact that programmer productivity has not increased substantially in the past 23 years is partially due to the pressure for more functionally complex systems and a programming environment that prevents most programmers from having direct contact with the computer (Ref. 11-17). However, the most prominant cause of low productivity is the lack of major advances in programmers' tools following the development of general compiler languages. The advances that are necessary to increase programmer productivity involve: 1) improving the methods for communicating computer system requirements between systems analysts and programmers and

2) replacing the standard one-dimensional, character-string oriented computer I/O with the direct approach of two-dimensional, interactive graphic I/O. The second represents an efficient way for achieving the first.

11.2.3 Trends and Projected Developments in Languages

11.2.3.1 General Purpose Languages and DOD Standardization Efforts - The largest improvement in programmer productivity occurred between 1955 and 1960 (Ref. 11-17), and this effort can be partially attributed to the development of compilers, particularly FORTRAN in 1956. The negligible increase in productivity from 1960 to 1975 can be attributed to causes outside the area of languages, such as the availability of time-sharing in 1963, which improved the programmer environment, and/or the improvements in hardware, such as faster computers and cheaper memory. However, there is no evidence indicating significant differences within general-purpose languages over this same period, and it is therefore concluded that very little (if any) of the improvement in productivity is due to the past developments in general-purpose languages.

It is highly unlikely that any breakthroughs in language design will occur in the next 2 years given that very little has occurred during the past 18. Concurrent PASCAL languages that facilitate parallel algorithms (e.g., PFOR - FORTRAN for PEPE) will help solve concurrent problems (operating systems) and parallel algorithms for special problems (weather processing) on special architectures. However, these improvements will not substantially affect productivity. Therefore, as long as the developments in general-purpose languages continue to parallel the developments over the past 23 years in terms of major breakthroughs, nothing new in this area of software production tools will improve programmer productivity in the 1980-1985 period.

The following six general-purpose languages will continue to dominate programming efforts and programmer productivity, primarily because of the inertia of the computer industry and the call for a particular standardization of lauguages within Government agencies:

- FORTRAN
- COBOL
- JOVIAL — Air Force
- TACPOL (PL1) — Army
- CMS-2 (FORTRAN) — Navy
- SPL-1 (PASCAL) — Navy.

The DOD standardization of these six languages will provide some advantages in terms of portability of software from installation to installation. In addition, there is an effort to develop a single DOD language (DOD-1) through the standardization of features associated with the above languages for "embedded computer applications" (Ref. 11-14). Unless the DOD-1 language addresses the major problem of specifying the requirements for computer-based systems in a comprehensive, explicit, and unambiguous manner, DOD-1 will not do much more than ensure the portability of developed software.

Other languages in this same classification that have been used in the past and will continue to be used during and after the 1980-1985 period are:

- PL1
- ALGOL
- PASCAL.

The basic difficulty associated with the development of computer-based systems, as stated previously, is the definition and design of system requirements. These requirements must be stated in terms that can be used to communicate and analyze system concepts before the system has been built. These requirements must be stated in terms that can also be used to test the system to ensure that it does what it is supposed to do after it has been built.

It is obvious that the above list of general-purpose languages does not satisfy those activities that must be performed before implementation of the system. It is also apparent that these languages cannot be used to test the system after it has been implemented without violating what is equivalent to the circular fallacy in formal logic (i.e., program listings cannot be used as criteria against which one tests the programs correctness). These languages do provide after-the-fact descriptions of the software. The descriptions become important when used in conjunction with system requirements that have been developed independently at a time prior to the implementation of the software. Because the basic difficulties

in system definition and design are not solved by the above languages, the languages will continue to have a negligible effect on the improvement of programmer productivity during the 1980-1985 period.

11.2.3.2 Simulation Languages - Simulation languages during the 1980-1985 period will continue to be used in the analysis process. However, they will continue to be too costly, too late, too inflexible to changes, never in sync with requirements, and never evaluated in an objective manner. These languages will continue to be used to model a system in a limited manner, almost always attacking only the interesting, academically elegant, sophisticated, new, and obvious parts of the system, leaving the major percentage of system design problems to be handled after the system has been built. This is essentially not the fault of the simulation languages as much as it is a fault in the way they are used and abused. Some exceptions in which the language is at fault are exemplified by SCERT, which requires that a primitive representation of the system requirements be generated (quite laboriously) on "cards" for input to the simulation run (Ref. 11-20). Although it provides a record of the system assumptions, it is very difficult to connect these with the requirements for the system and therefore leads one to be suspicious of the SCERT run results (GIGO). To compound the problem, SCERT simulation results are supposed to be used to configure the hardware at a time when not very much is known about the system functions, thus forcing the systems analysts to fit the requirements to the hardware.

There is a close relationship between simulation languages and system requirements specification/analysis languages in that both representations of a system are used with the objective of analyzing the system through models, the major difference being that the latter should be totally comprehensive and complete in describing the system. As a result of this close relationship, simulation languages will be quite useful in suggesting solutions to the problems encountered in system requirements specification/analysis language development. The major contributors will continue to be:

- GPSS
- SIMSCRIPT
- SALSIM
- SIMULA.

11.2.3.3 Graphic Languages - Graphic languages are obviously dependent on graphic hardware developments. This means that the use of graphic languages and graphic techniques for computer I/O has been limited by the cost of graphic hardware and therefore its availability in the past. Currently, progress is being made in decreasing this cost in all areas of graphic equipment (see Section 10). Examples of this are: TEKTRONIX 2000 series storage tube display terminals, DEC GT-40 series refresh type display terminals with light pen capability, and VARIAN and VERSATEC printer/plotters, all at prices under $50,000, making interactive graphics a reasonable approach to the man-machine computer interface problem.

Assuming that progress will continue to be made in the area of graphic hardware costs and availability, graphic languages will be very effective in solving a portion of the productivity problem by providing both programmers and systems analysts with a higher level of communication with the computer systems they use and build. These graphic interfaces will eventually replace the one-dimensional, string-oriented (for card reader or textual data displays) interface languages, which pose too many nonhuman, nonproblem associated restrictions.

Progress in the area of standardizing graphics language facilities began in the 1960s when Bell Laboratories developed GRAFPAC (Ref. 11-21). The objectives of GRAFPAC were to unify the descriptions of graphic data and provide graphic output on a number of different devices without the need to regenerate the data. For example, GRAFPAC was capable of drawing lines and circular arcs, plotting points, and generating text on devices such as SD-4060, FR80, and ADP-5000 microfilm recorders; the Calcomp plotters; Gerber; STARE; and the GRAPHIC 101A (Ref. 11-21). Automatic and direct scaling as well as "windowing" were provided to give the GRAFPAC language user the ability to select and draw portions of an overall graph or drawing.

This effort is continuing as evidenced by the results of a recently published report of the "Graphic Standards Planning Committee of ACM/ SIGGRAPH" (Ref 11-13). The overall objectives of the committee report were to provide a standard that would ensure programmer portability as well as program portability. Also, the report specifies the requirements for a "core" graphics system with a user interface that is device-, machine-, and operating system-independent in addition to being usable from high-level application languages such as FORTRAN and PL/1. The report included a survey of the following graphics packages:

- ADAGE
- CALCOMP
- DISSPLA
- GCS
- GINO-F
- GPGS
- IG
- TEKTRONIX.

The report noted the following general features which differed among the packages: "A general disarray in graphical input, device coordinate systems differ considerably, different levels of usage of picture segments, non-uniform support of positioning and drawing (e.g., 2D and 3D, relative and absolute, and polyline), several different approaches to textual output, and differences in quality and handling of error reporting." If the Graphics Standards Planning Committee meets its objectives, all of these problems will be solved, much to the advantage of the graphics language user.

11.2.3.4 Formal Requirements Languages - Formal requirements languages combined with interactive graphics should improve programmer productivity in the 1980-1985 period. During the late 1960's, the computer industry as a whole started to become aware of the importance of specifying computer systems in detail before they were built. This was beautifully expressed in a paper presented the 1969 Spring Joint Computer Conference (Ref. 11-22):

> Of all the problems facing the developer the lack of an adequate systems specification is both the hardest to overcome and at the same time the one which, if not overcome, will cause the most grief. This specification is a statement of what the system will do under all conditions, how it will do it, how fast it will operate, and how large it will be.
>
> From a practical point of view, however, one can regard it as axiomatic that what is not contained in the specification will not be in the system when it is finally built.

In spite of these statements as well as many others written since 1969 advocating that we <u>think</u> out our computer systems before we build them, and in spite of the overwhelming evidence of poorly built computer systems, the industry continues to avoid specifying computer systems (before they are built), usually with such obscure objectives as "maintaining design freedom until the last moment". This somewhat contradictory position is exemplified by the following quotations taken from Reference 11-15:

> Generally speaking, previous experience with large-scale software development has been depressing. The symptoms of the inadequacy of our software design and development methodology are high costs, unresponsive products, slippage of production schedules, and difficulties in system operation and maintenance. ... The primary objective of a specification methodology is to rectify the above problems. To accomplish this objective, the methodology should provide means to <u>precisely state the system requirements</u> and provide analytic procedures to check for the consistency, completeness, and correctness of the specification.

While in the same report, these objectives are compromised by the following statements:

> Another important concept that has been developed recently in the area of system design is the family of systems approach. The essential concept is that we are developing a system conceptually in a hierarchy of levels, we are starting out with a broad class of systems, and as we go along and make design decisions, the class of systems is limited to meet more specific requirements. If a

> system is designed with this philosophy, requirements change and, in general, system modifications can be accommodated much more easily by zeroing in the proper level at which a design decision must be changed. ... We have made an assumption that design freedom should be maintained as far as possible.

This philosophy implies two faulty assumptions: The first is that design creativity is best nurtured in an atmosphere of total design freedom. This same premise has been widely used as an argument during the sixties and seventies for inhibiting the use of precise, unambiguous system requirements specifications at the upper levels of computer system definition and design. However, there are some extremely creative people who disagree with this. For example, Igor Stravinsky states (Ref. 11-23):

> What delivers me from the anguish into which an unrestricted freedom plunges me is the fact that I am always able to turn immediately to the concrete things that are here in question. I have no use for theoretic freedom. Let me have something finite, definite -- matter that can lend itself to my operation only insofar as it is commensurate with my possibilities. -- Well, in art as in everything else, one can build only upon a resisting foundation: whatever constantly gives way to pressure constantly renders movement impossible. My freedom thus consists in my moving about within the narrow frame that I have assigned myself for each of my undertakings.

The second faulty assumption is that there is only one "family of systems" when in fact there are two: the "family of requirement systems" and the "family of solution systems". Limiting the class of requirement systems is "definition", while limiting the class of solution systems is "design". The former activity is the formulation of a problem in terms of precise, unambiguous system requirements that state what the system is to do, when, and how fast. The latter activity is the orderly division of these system requirements (commonly referred to as "decomposition" or "allocation") into units (at the lowest level) that have technologically implementable solutions. Until the system requirements have been explicitly

stated, there are no criteria by which one can limit the class of solution systems (i.e., the family of solution systems remains infinite).

In practice, the industry avoids the definition activity (as described above) and thus each analyst responsible for units at the lowest level of design is forced to develop his own independent criteria. These criteria are usually either lost or buried in the solution documentation (program listings). When the system is finally integrated and tested against another set of independently determined test criteria, we wonder why it doesn't work.

The conclusions are:

- There are two distinct families of systems: the family of requirement systems and the family of solution systems.

- Formal, precise, unambiguous system requirements provide a necessary service by bounding the family of solution systems and therefore these same requirements facilitate the search for a set of "best" solutions.

In studying the past and present system requirements specification/analysis languages, the following 19 languages were identified:

- ADS (Ref. 11-15)
- BDL (Ref. 11-15)
- FSM (Ref. 11-15)
- F^2D^2 (Ref. 11-15)
- HIPO (Refs. 11-15 and 11-24)
- HOS (Refs. 11-15, 11-25, and 11-26)
- IORL (Refs. 11-27, 11-28, 11-29, and 11-30)
- LOGOS (Ref. 11-15)
- PDL (CSC) (Ref. 11-31)
- PDL (CFG) (Ref. 11-32)
- PSL (Refs. 11-15, 11-26, 11-33, and 11-34)
- RSL (Refs. 11-15 and 11-35)
- SADT (Refs. 11-15 and 11-36)
- SDDL (Ref. 11-37)
- SODA (Ref. 11-15)
- SSL (Ref. 11-38)
- TAG (Ref. 11-15)
- THREADS (Ref. 11-15)
- VG (Ref. 11-15).

In almost all cases, each of these languages suffers from one or more of the following deficiencies:

- No formal syntax and/or semantics
- Most are imprecise, ambiguous, and textual in format and therefore indeterminate and nonquantifiable.
- Current languages tend to concentrate on highly developed hierarchical structures for containing mostly ambiguous and undiscernable prose.
- Descriptions are limited to only parts or levels of the system description, which defies unified and comprehensive system analysis, validation from level to level, and/or requirements tracing.
- Although several claim to be graphic type languages, all of those that are computerized except two (IORL and RSL) require noninteractive, one-dimensional, character-string input to generate the graphics.
- Of those that address timing, most treat it as an afterthought which is not integrated into the requirements language.

The results of these deficiencies have been and still are that one cannot see what the system is supposed to do using most of the currently available system requirement specification languages until the system has been implemented and all the money has been spent. This, of course, defeats the objectives for a System Requirements Specification/Analysis Language. There are exceptions to these SRSAL failings, but generally they all fail one or more of the above points.

The exceptions are RSL and IORL. In studying these two languages it is interesting to note their similarities. For example, with the exception of the "validation nodes" used to specify timing in RSL, the RSL "R-NET" diagram is almost identical to the IORL "IORTD", even to the extent that the symbol shapes for a given semantic definition are identical. It might appear that the two languages were designed by the same person or persons.

A noticeable difference between these two languages is that RSL seems to be limited to the area of "Control Dominant Software Architecture Specification" (Ref. 11-13), while IORL was designed with the objective of providing all levels of computer system specification from system definition to detailed software design.

The outlook for SRSALs in the 1980-1985 period is dependent on whether the computer industry begins to recognize the need to specify systems precisely and unambiguously before they are built. It is also dependent on the amount of research support (dollars) that are directed toward the development of SRSALs. The technology is here (i.e., interactive graphics and formally defined SRSAL). If we are extremely optimistic, these languages will be available and will have the following characteristics:

- Single language for all levels of system definition and design defined through formal syntax rules and semantic definitions (closure property).
- Language elements designed such that it will formally connect the total set of system requirements in a single hierarchy (library).
- Machine readable from an interactive graphic terminal, thus providing rapid modification of system requirements.
- Provide an extremely high level of analyst interaction through two-dimensional representations of all mathematical forms and logical conditions. Interactive graphics will be the primary I/O vehicle.

These languages will be capable of providing the following services directly from their representations of the system requirements and the use of interactive graphics:

- Provide fast access and modification of requirements information
- Provide simulation results derived directly from the system requirements
- Provide computerized validation of requirements at each step in the system development

- Provide computerized tracing of requirements from level to level within the system requirements hierarchy

- Each set of requirements at a given level within the system requirements hierarchy will provide requirements for the next level below and at the same time will provide integration instructions and test criteria for the system integration phase of development.

- Provide the structure for automated configuration management

- Provide management with specific quantifiable information regarding progress of the system development

- Provide the information necessary to effect changes to the system

- Provide a means for computerized evaluating of alternative solutions to a given set of system requirements.

11.2.4 Summary of Conclusions

In spite of the proliferation of programming languages and software development techniques, these software tools have not been effective in increasing the productivity of programmers over the past 23 years.

Because basic difficulties in system definition and design are not solved by the above general-purpose languages, they will continue to have a negligible effect on the improvement of programmer productivity during the 1980-1985 period.

Assuming that progress will continue to be made in the area of graphic hardware costs and availability, graphic languages will be very effective in solving a portion of the productivity problem by providing both programmers and systems analysts with a higher level of communication with the computer systems they use and build.

Formal requirements languages used in conjunction with interactive graphics offer the only hope for improved programmer productivity in the 1980-1985 period. However, these will not come to the rescue unless the computer industry is able to recognize soon that programmer productivity is a system requirement specification and analysis problem.

11.3 SOFTWARE METHODOLOGY

Software methodology is the practice of the method by which software is conceived, produced, and maintained over its life cycle. Although software has been produced for over 25 years, little attention had been given to software methodology until the last decade.

Software costs are personnel costs, and unlike hardware costs, which benefited by technological advances and steadily decreased for individual components, software costs have risen steadily as salaries have risen. Today, software methodology is undergoing close scrutiny in an attempt to define and refine it with a goal to reduce the cost through increased programmer productivity. Strong pressure is being applied by the user community for improved software.

This section analyzes the state of the art and the trends and forecasts the methodology for 1985. Normally, the state of the art refers to the technological limits that a science can be pushed in the present environment. However, due to the lack of metrics in software, there are no "leading" methodologies because no methodology has been scientifically proven to be better than others.

Since the analysis of software methodology is primarily the analysis of the way people practice it, in this section the state of the art is presented as the way the majority practice it. Newer methods that are used by a minority of people (or organizations) are presented primarily as trends. Finally, using a model of the methodology of a related field, modified by the current state of the art of software practice and its trends, a forecast is derived through 1985.

11.3.1 State of the Art of Software Methodology

This section is divided into four subsections. The first describes the state of the art of software methodology in general as it applies to all phases of the software life cycle. The final three subsections describe the state of the art as it pertains to each of the three phases of the life cycle.

Throughout these four subsections, the "state of the art of software methodology" refers to the application of software methodologies as generally practiced by the scientific applications programming community today.

11.3.1.1 The General Status of the State of the Art - Software methodology today is a product of the experiences of the past. In order to document the present and to develop the reasons for the use of the methodology of the present, it is necessary to know and understand the past. The subsection first describes the history of software methodology before describing the state of the art.

11.3.1.1.1 Background - Software methodology is little more than 25 years old. The first computers were functionally slow, limited in memory, and simple. Programs were limited not only by the computer but by the tools available for program creation and "debugging". The first programs were written directly in machine language (literally 1's and 0's), and debugging consisted of testing against hand-generated test cases. The Initiation Phase consisted of no more than the statement of the problem to be solved. Since most early programs were designed, coded, and tested by the same programmer, user and programmer interactions were straightforward and clarifications could be made whenever the programmer or user felt necessary. Programmers developed their own methodology and adjusted it to their own personalities. The Development Phase ended when the programmer announced that he was done. There were little necessity for a formal test, as maintenance was performed by the same programmer. Rarely was a program turned over to the end user, but remained with the programmer who ran it on request.

In reality, programmers were more concerned with the speed and efficiency (in terms of memory requirements) of their programs and in interfacing directly with the computer (operating systems did not exist) than with methodology. Since the computers were used primarily to solve problems that could not be solved without them, and since software costs represented only 25% of data processing costs, little attention was given to methodology or programmer productivity.

From this humble start in the early 1950's, the computer industry underwent an explosive growth. Computers increased rapidly in both speed and storage capabilities. No longer was the computer used only for special problem solving, but extensively for business and scientific applications. Programs of increasing size and complexity were conceived and built.

As programs became bigger in the late 1950's, groups of programmers were teamed together to build them. No one gave serious thought to the exponential rise in programmer interactions required by groups of programmers working on the same program or to the methodology they used. Success or failure of a software effort was considered a problem of organization and management, not methodology. The small program techniques continued to be used but in slightly more organized form. Programmers were given pieces of the larger program and then proceeded to use the same informal design, coding, and testing techniques that they had used in the early fifties. They were also given a schedule, based on the overall schedule, but project schedules were optimistic, and because of the relative isolation of each programmer, schedules generally collapsed when various pieces of the program were integrated together. Yet software, in spite of failures, had some notable successes.

The first successful large program developed was the SAGE system, a complex system to defend against manned bombers, completed in the late fifties. This system was conceived and prototyped by a small group of programmers using a machine with only 256 words of memory. The actual system was to have considerably more capability than the prototype;

the computer was to have 8,000 words of memory. Within a year, 1,000 people were involved in the development. The system computer expanded to 64,000 words. The original requirements were cut considerably and the system was delivered late and well over budget. The system did work, however, and thus was termed a "success". Few people looked at the negative side of the project, as the introduction of higher-level languages (FORTRAN in 1956 and COBOL in 1961) and system software were widely accepted as the solution to problems such as encountered by SAGE.

Buoyed by the successes and ignoring the failures, industry and Government alike looked to the computer to solve complex problems and automate manual processes. In the early 1960's, the business community made concreted efforts to produce programs to handle payrolls, inventories, and other accounting functions, while Government, particularly the military and space programs, became totally dependent on them. As machines increased in capability, more demand was placed on the software to produce the increasingly elaborate programs conceived to utilize the capability. Throughout the early and middle sixties, the number of programmers grew but the methodology they employed did not. The only change in methodology was the slow acceptance of higher-level languages, but programmers generally still preferred assembly languages.

By 1968, 2 years after the IBM 360 was introduced, the impact of a new generation of computers with an order of magnitude improvement over their predecessors was beginning to be felt. It is estimated (Ref. 11-41) that computer performance/cost had increased by a factor of over 100 (normalized to 1955), while individual programmer productivity increased by a factor of only 2.2. *Software costs now equaled hardware costs; project cost overruns and late deliveries were commonplace. Most everyone agreed that software was a "problem". What was visible, however, was not the problem but the effects of the problem. The industry did not know the cause of the problem. Managers in the sixties had applied their best techniques **and increased their controls, but it still was a basic rule

*The increase was due to the use of systems software and higher-order languages.

**It should be noted that most software managers were not programmers but engineers and business administrators. The management techniques that they used are not suited to software management, nor did these managers have the experience to recognize the causes of the problem.

that "the larger the project, the larger the disaster". No matter how well-planned and managed, large projects seemed to "take off on their own" and were soon out of control. Yet they were always on schedule until integration started, when they seemed to stand still. (This is a form of what is called the 90% syndrome in software.)

In 1968, a significant but little-noticed event took place. The first Software Engineering Conference (sponsored by NATO) was held. The significance lies in the fact that for the first time formal recognition was given to the idea that programming is not purely a science nor is it a craft (as programmers practiced it), but instead it was recognized that programming should be practiced as an engineering discipline. The problem was recognized as software methodology!

The following summary of software methodology as practiced in 1968 is presented as a baseline for comparison throughout the rest or this section:

Initiation Phase - Methodology in the Initiation Phase could be characterized by informality and inconsistency:

- The user conceived a requirement for a program on an existing computer* or for a hardware/software system (i.e., a radar).
- The user (or in the case of a system, a group of his representatives) attempted to define the requirements of the program or system.
- In the case of a hardware/software system, feasibility studies concentrated on the hardware, as did the definition of requirements. For the case of a program, feasibility studies were omitted.
- Simulation as a tool for system or program feasibility was nonexistent.
- The output of this phase almost always was a document written in English with the following characteristics:

*Such as a new accounting program or a new data reduction program.

- Lack of standards or formats. It informally described the requirements in prose.

- Lack of a description of intended operation. The operational document was left for the programmer to write after project completion.

- Lack of specific acceptance test requirements.

- Requirements for software characterized by:

  - Inconsistent or impractical requirements
  - Ambiguous or unclear requirements
  - Lack of critical requirements such as timing.

- In the case of systems, hardware requirements sections were separated from software requirements sections, and the two were often in conflict.

- In the case of medium and large systems after completion of this phase, the originators of the requirements documents were reassigned to other tasks and unavailable to provide clarification.

Development Phase - The methodology used during this phase could be generally characterized as an organized form of the small program methodology that has evolved since the early fifties. It should be noted that the methodology always worked fairly well for small programs (less than 16,000 commands and requiring fewer than five programmers) where the number of programmers and program module interactions were small. The characterization that follows is representative of the development of medium and large programs and hardware/software systems. The problems described increase in magnitude exponentially with the number of interactions required.

The methodology of the Development Phase can be described as follows:

- Projects always started out with complete plans containing:

  - Milestones for requirements specifications documents, preliminary design reviews, etc. Development stages were well recognized and milestoned. Equal time spans were allocated for design, coding, and testing.

- Manning schedules were characterized by a bell-shaped curve with peak loading during the coding period.

- Organization charts featured a project manager, his technical group (the systems group), module group leaders, and programmers. Module group leaders were scheduled for the design stage, programmers for the final stages of designing and coding.

- Elaborate PERT charts outlining the "critical path".

- Concurrently with the planning phase, the requirements specification stage was started using the system group as a base. This group set about specifying the system (or program) based on their interpretation of the requirements. (If this was a hardware/software system, a similar hardware activity was proceeding independently in another area.) Unlike at the time of proposal when the basic philosophy was to minimize every software function to reduce costs (and if competitive, costs were probably cut more), the philosophy of the systems group was not to interpret every specification in the user's favor and to provide dual functions to handle conflicts in the specifications. It was too late to worry about feasibility, so no one did. It was too early to worry about timing and sizing requirements, so no one did, although there was a feeling that "things were tight". The document output by this effort could be characterized by:

  - All the characteristics of the original requirements document

  - A far more elaborate description of the functions to be performed than in the original design concept

  - A different (and normally larger) set of modules in the high-level program structure than were used for the initial concept.

- As the same people who performed the requirements stage were required to help with the planning and the period was originally scheduled to be brief, and the system turned out to be more elaborate than originally thought, this stage ended "slightly" behind schedule.

- Even before the requirements specification stage was completed, the design stage began (preliminary design) on time as scheduled. The group leaders (with a few extras "to help") arrived one by one as they completed their previous assignments. The qualifications required to be a group leader were:

- ▲ Any programming experience on another project
- ▲ At least 3 years' experience, although there were exceptions (this still was a young and rapidly growing field)
- ▲ Availability.

- Group leaders who arrived before the new specifications document was published designed their module from the original requirements document, while those who arrived after the new document was published used both. In either event, group leaders conceived their module based on their own personalities.

- If not enough group leaders arrived in time, members of the original systems group assumed their positions.

- The end product of the preliminary design required a design document containing an English description of the module, identification of major interfaces with other modules (although not detailed), and design of the module with at least the identification and description of the major routings (with flow charts). Design methodology was oriented to these ends. The description of the module was produced with a combination of reading, thinking, and interpretation of whatever initial documents were available. If the module had significant algorithmic content, it was generally unavailable (the algorithm developers in another area were still conceiving the algorithms) and ignored. Interfaces were settled verbally between group leaders on a gross level and then written down. The last step in the design was the hardest due to the unknown effect of the algorithms (and hardware), but the module was divided up into major routines based on the module leader's conception.

- During the preliminary design, what was left of the systems group was:
  - ▲ Working on a new proposal ("temporarily")
  - ▲ Reworking the requirements specification document, which had been reviewed by the new representatives of the user and was found lacking because:
    - It did not meet his new interpretation of the initial requirements document.
    - It was important to squeeze every last enhancement out of the software in order to obtain more software for the cost.

- Due to increased and unexpected complexity of the software, the lack of clarity of previous documents, and reiterations of the preliminary design due to changing requirements, this portion of the design stage always ended late, sometimes so late that a schedule slip had to be conceded and a new schedule developed. More often, however, the preliminary design was completed before the halfway point in the schedule for the detailed portion of the design, so it was agreed that the schedule could be made up. It should be noted that another technique to handle the apparent overrun at this point was to terminate the preliminary design, saving what was originally planned for the detailed design.

- Preliminary design documents could be characterized by all the attributes of previous documents with more elaboration.

- When the preliminary design review was held, normally long after detailed designing had started, the following groups criticized the design:

  - The user's representatives for the same reasons that he criticized the requirements specification

  - The algorithm designers because it didn't appear that the modules as described matched their preliminary designs

  - The hardware designers for the same reasons as the algorithm designers.

  After the design review, since the project was behind schedule, it was normally agreed to incorporate all criticisms and changes into the detailed design.

- The detailed design started on schedule or as near to on schedule as possible when additional personnel became available to help with the schedule slip (or because it was supposed to be time to code). The detailed design effort can be characterized by

  - A repeat of the preliminary design effort but on a bigger and more detailed scale

  - Mounds of documentation produced using the design methods described before* to the lowest level in each module

*The order of breakdown was one of convenience; low-level routines might well be designed first so that they could be coded early.

- Generally detailed attempts to define as many internal module interfaces as possible (the algorithms were still missing)

- A slow realization that someone should estimate file size and memory requirements. Timing requirements were not forgotten, but no one knew how to determine whether the requirements could be met.

- The detailed design phase formally ended (although informally it would continue to the end of the project) with a critical design review and a set of design documents but often still lacking the final versions of the algorithms.

- The detailed design review produced new criticisms and suggested changes by the same groups that attended the preliminary design review for the same reasons as at the time of the preliminary design review. The detailed design review caused a limited re-issue* of the design documents with the suggested modifications.

- The coding and integration phase** started on schedule or as near to schedule as there was something to be coded. The technique used is today called "bottom-up" coding, since the lowest identified routines were generally coded first. Each module was generally constructed step by step, starting with the lowest identified routines and gradually working upward to the "driving" routines. The coding process went as follows:

  - A programmer was assigned a subroutine to code by his group leader. This routine was generally a routine from the lowest level left to code for which the group leader felt sufficient information could be obtained (i.e., the algorithms could be designed by the programmer or obtained by contacting the algorithm group).

  - The programmer studied the documentation (hopefully the last version), which might consist of a short description and a macro flow chart†.

  - The programmer obtained and studied any required algorithms or derived his own.

*There were now two sets in use and still not enough to go around!

**In 1968, these were considered more as separate stages because of the methodology.

†Macro flow charts seldom described the code and were more often functional flows that included critical decision branches.

- After deciding that he understood the problems, the programmer organized the program (optional) into a driver and still lower-level routines. The programmer produced (optional) more detailed flow charts on scrap paper.

- When the unit was designed, the programmer looked around (optional) to see if he could find any one else who had routines he could adapt to meet his requirements.

- For each routine in the unit, generally starting with the lowest level:

  - Assembly or a higher-order language was chosen for coding.

  - The routine was coded and submitted for compilation and assembly.

  - After submission, the programmer decided on his test procedure and designed and coded his test routines.

  - When all routines were compiled (assembled) without errors, they were married together and resubmitted to start the tests.

  - After each run, the results were analyzed to determine if the routine had performed as expected. If not, both test routine and routine under test were studied and corrected and then resubmitted.

- When each piece of the unit was completed to the programmer's satisfaction, it was combined with other pieces, often by the lead programmer, a new test was designed, and the whole assembly was retested. If the test failed, the faulty program(s) was modified (or "patched") and the test repeated. This activity is called integration.

- Sometime during the coding and integration phase, some or all of the following happened:

  - The manager was told by his staff and told the user that the software was 90% finished.

  - The algorithm designers discovered mistakes and better algorithms and asked that changes be made in the software.

- The programmers discovered that some of the algorithms could not be implemented in the software and asked the algorithm designers to redesign the algorithms.

- The computer reached 100% utilization (no one thought about the computer requirements for the project until that point).

- The workload submitted to the computer reached 120% (or more) of the computer's capability.

- An integration test showed major interface problems requiring extensive redesign of one or more of the modules.

- An integration test failed because the program did not fit in core, resulting in the need to rework the code (often recoding routines in assembly) or a request for more memory.

- The best code counts the manager could obtain told him that more than twice the amount of code had been generated than any of his worst-case estimates. (No one until this point had thought to include test and support code in the estimate.)

- The scheduled project end-date passed.

- An overrun was declared, resulting in the following actions:

  - The manager was replaced with a new manager (optional).
  - A new schedule was developed.
  - The project was recognized as in the first step.
  - The user was asked for more money.

- If more money was obtained, the above steps were repeated, except that problems described with the algorithms became problems in interfacing with the hardware.

- Somewhere along the way, someone developed acceptance tests much in the fashion of the rest of the software and obtained the user's approval. Although the tests were run many times, projects usually ended when:

  - The user decided to accept anything and finish it himself.

  - The user changed his requirements (including timing) so that the last run of the acceptance test could be declared acceptable.

- ▲ The user ran out of money (some corporations in the sixties actually went bankrupt because of their data processing systems!).
- ▲ The system no longer was required or necessary.

Although this description of a 1968 project is a worst-case composite scenario of software development, most projects during the late sixties experienced all or some of these problems to a degree related to the size and complexity of the program under development. Virtually all software produced during this period had the following characteristics:

- It was filled with errors (testing was incomplete).
- It was ragged. There were no standards for module or subroutine size or organization; code styles varied, commentary was criptic; and documentation was out of data or incomplete.

The methodology used to produce the program had the following characteristics:

- An attempt at good management at the top but virtually no management at lower levels of the software organization
- Milestones that could not be measured
- Lack of management tools for project monitoring, such as code monitors, cost estimation, and timing and sizing.
- Generally forced, small programming methodology
- Lack of programming tools
- Lack of systems techniques or philosophy.

Maintenance Phase - The Maintenance Phase started out as a continuation of the Development Phase since testing was seldom fully completed and the system failed to meet its requirements fully or the requirements had changed. As a rule of thumb, 10% of the development staff was left behind to perform the maintenance, and on large systems they were joined by new programmers directly employed by the user.

The Maintenance Phase was characterized by:

- Long periods of study to determine how the program actually worked. This was required due to the inadequate documentation delivered and often resulted in the first accurate flow charts for the routines. These were kept by individual programmers (job security) who became "experts" for those routines.
- Panic design and patch sessions to correct the errors that caused system "crashes"
- Lack of formal configuration control
- Lack of any formal attempts to maintain or improve the existing documentation
- A general feeling that the software should be continually improved
- Little or no testing of new or patch code
- Panic redesign and patch sessions to correct the errors caused by the code that was added to fix the original errors
- Major efforts to enhance the software, with all the characteristics applied to the Development Phase although less organized.

In spite of the problems throughout the software life cycle, large software systems in both the commercial and scientific world came on-line in the sixties and they worked. The software may not have ended up with all the features the user had envisioned and may have cost more than he expected, but when it finally became operational it generally represented a major increase in the user's capability or payed for itself (in the case of automation) by reducing costs. Medium-sized programs (16,000 to 64,000 lines of code) generally became highly reliable early in the Maintenance Phase. Although people wondered if there were a limit to the size of the software program that could be developed, the software industry was willing to develop software for any application as long as it was not to be done with a fixed price contract.

Software methodology during the early seventies showed little change from the methodology of the late sixties although the demands on it continued to increase. Computers continued to become more powerful and the applications more complex. The trend toward more management control increased (even as it had in the sixties). The general software industry was unaware of software engineering of new methodologies.

The seeds of change had been planted in 1968, however, and research into engineering-oriented methodologies was in process. Some of the early research in methodology resulted in the following:

- Early cost estimating models - Early cost estimating models depending on estimating the number of commands in the final program but recognizing and predicting the costs due to other factors such as programmer interactions, support code, and documentation requirements.

- Proposed criteria to be used in decomposing systems into modules (Ref. 11-42)* - Modular programming is a commonly accepted method of designing systems. Conventionally, a system is decomposed into modules based on flowcharts with proper interfaces that provide flexibility, comprehensibility, and manageability of the program. This conventional approach has severe drawbacks. For example, if the specifications are changed, many modules may access a few data structures, thus hindering swift modification. A new set of criteria is proposed to eliminate these drawbacks while still maintaining the benefits of the modular programming. The criteria are:

  - A data structure, its internal linkings, accessing procedures, and modifying procedures should be part of a single module.

  - The sequence of instructions necessary to call a given routine and the routine call itself should be part of the same module.

  - The formats of control blocks used in queues in operating systems and similar programs must be hidden within a control block module.

  - Character codes, alphabetic orderings, and similar data should be hidden in single modules for greatest flexibility.

*Although the research is attributed to the original author, this description is taken from Reference 11-43.

- The sequence in which certain items will be processed should be hidden within a single module.

The new set of criteria will, in general, generate more modules than the conventional method merely because the functions are more refined. This will cause greater procedural call overhead than in conventional programs. To save the call overhead yet gain the same advantages of modular programming, the implementation of the modules must be in an unusual way. The routines could best be inserted into the code by an assembler or by using highly specialized and efficient transfers for procedural calls. The proposed way of decomposing a program is to gather a list of design decisions that are difficult or likely to change. Each module is designed to hide such a decision from others. Since in most cases design decisions transcend time of execution, modules will not correspond to steps in the processing. To achieve an efficient implementation, one must abandon the assumption that a module is one or more subroutines and instead allow subroutines and programs to be assembled collections of code from various modules. Other early research in methodology resulted in:

- Stepwise refinement (Ref. 11-44)* - Stepwise refinement is the process of designing and constructing a program by gradual development of partial solutions. In each step, one or several instructions of the given program are decomposed into more detailed instructions. Some well-known techniques such as strategy of preselection, stepwise construction of trial solutions, introduction of auxiliary data, and recursion are used. In this process of successive specifications refinement, decisions concerning details of representation are deferred as long as possible. Each refinement in the description of a task is accompanied by a refinement of the description of the data that constitute the means of communication between subtasks.

- Egoless programming (Ref. 11-45)* - Conventionally, programmers identify programming errors with programming incompetency. As a result, programmers are defensive about their programs. Due to this natural phenomenon, many extra difficulties are created during system integration tests. From the psychological point of view, Weinberg (Ref. 11-45) proposes that a new school of thought be brought into the programming profession:

*Although the research is attributed to the original author, the description is taken from Reference 11-43.

programmers should disassociate their egos from their programming mistakes. This technique can provide better working relationships among team members and higher production for each individual. The concept of the "walk-through" came from this book.

- Structured form (Ref. 11-46) - This paper contains the theorem and proof that any program can be written using three basic control constructures: Process, If...Then... Else, and Do...While. It also introduced the rule that each function should have a single entry and a single exit.

- Structured programming - Dijkstra (Ref. 11-47) proposes a methodology to construct the designer's mechanism in such a way (i.e., so effectively structured) that at every stage of the testing procedure the number of relevant test cases will be so small that he can try them all. It is therefore a method of structuring the program so that it can be "exhaustively" tested and confidently verified.

Most of this early research focused on particular aspects of software methodology. However, the individual methodologies were combined in a practical manner and applied to the development of the New York Times information bank. The methodology (Ref. 11-48) combined the philosophies of structured programming and structured form with a management philosophy that utilized chief programmer teams. The nucleus of a chief programmer team (Ref. 11-49) consists of:

- A chief programmer - a senior-level key programmer who produces the critical nucleus of the system in full and integrates all the other programming activities

- A backup programmer - a senior programmer who works closely with the chief programmer

- A program librarian - a programmer/technician who maintains a program library containing test data and the current baseline version of the program.

This approach (also called top-down or structured design, coding, and testing) utilized a rigorous step-by-step process of program design (top-down design) in which functional specifications are expanded in a top-down manner into simpler and simpler subfunctions until statements of the

programming language itself are reached. The design at each step is converted into code directly using the structured form of coding.*

The project, which claimed an almost error-free end product and a productivity rate of 35 lines of code per day, attracted wide attention, and the industry in general began to look at software methodology. In 1975, the first annual conference on software engineering was held in the United States (under the auspices of the IEEE). This conference further focused attention on software methodology. Finally, in 1976, the Department of Defense released DOD Directive 5000.29, "Defense Systems Software Management Plan", and with it money to support research in software methodology. The industry was sure that the software "problem" was one of methodology.

11.3.1.1.2 State of the Art - In determining the state of the art of software methodology, the following facts stand out:

- Most of the research in software methodology has occurred in the last 7 years.
- The computer performance/cost ratio is still increasing by a factor of 10 every 5 years, implying more demands on the software (Ref. 11-41).
- Programmer productivity has increased by a rate of 3% per year during the last decade (Ref. 11-41).
- The motivation of improving software methodology is based partially on the beliefs that software costs are increasing faster than hardware costs and that programmer productivity can be increased. Both of these beliefs are currently under attack as myths.
- Software managers are concerned with "saving" projects in trouble.
- The idea that programming is an engineering discipline is becoming common.

Most of the facts need no explanation as they are evident from the history presented in the previous section. The fourth fact, however, is different. The belief that costs were rising is attributed to Boehm

*As a result of this project, the term "structured programming" became associated with the methodology of the project.

(Ref. 11-50) and others who have predicted that software costs will be three times hardware costs in 1978 and will approach 90% by 1985. It is this fear of rising costs that led to the DOD Directive in 1976. The solution to rising costs is seen in the industry to be increased productivity. The attack on these beliefs is lead by Frank (Ref. 11-51). He contends that the idea that programmer productivity can be significantly increased is a myth. He feels that there are currently only two important ways to increase programmer productivity: higher-level languages and control of software production. The former has had no significant impact since the introduction of FORTRAN and COBAL in the late fifties and early sixties, and today (Ref. 11-41), "the industry no longer believes that there is any magic in higher-level languages, or in operating systems ... (although) it still continues to create them and call them magic".

Control of software production seems to be the only hope for increasing productivity, but as Frank again points out: "Different organizations and different philosophies of operation can produce equal results". Stated another way, as far as can be ascertained without software metrics, no one approach to software production is superior to any other.

On the question of software costs, he argues that both Datamation and International Data Corporation (IDC) annual cost surveys show that software costs have remained equal to hardware costs over the last few years and that predictions that software costs will reach 80 to 90% of hardware/software systems costs are wrong. There are three reasons for this:

- Hardware money saved on cheaper computers is spent on more and better peripheral equipment (terminals, timesharing, memory, etc.) to support the increasingly complex applications.
- The availability of package software is growing and software users are increasingly considering its use, thus spreading its cost over many users.

- Cost trends such as Boehms's include systems software. However, in applications software, systems software is an extension of the hardware and its cost should be spread over many computers and amortized over the life of the computer system. Of course this argument breaks down in special-purpose hardware/software systems such as required for military purposes when the systems software must be replaced.

In assessing the state of the art, Schwartz's commentary (Ref. 11-52), after reviewing a series of software projects over the last 20 years, is striking: "this and numerous other examples seem to illustrate that things haven't changed very much in the past 20 years in carrying out programming projects". Equally striking is Dolota's assessment (Ref. 11-41): "Programming is still a craft." Certainly the relatively small rise in programmer productivity lends further credence to this view. However, when combined with the other facts a more complex view emerges. The general state of the art of software methodology* can be summarized as follows:

- The basic methodologies used (showing pictorially in Figure 11.3.1.1.2-1) greatly resemble the methodologies of 1968 but are in the process of change. (The changes are discussed as trends.)

- The industry understands what goes wrong and why during the course of a project but is unable to prevent the problem from occurring.

- The importance of applying engineering techniques to the Development Phase is well recognized. However, this application to date has taken the form of improvement of existing techniques (1968) rather than the adoption of newer methodologies.

- Software engineering is primarily limited to the Development Phase of the life cycle and has yet to develop a systems viewpoint.

These general statements will be amplified in the following sections.

11.3.1.2 State of the Art as It Relates to the Initiation Phase - The Initiation Phase until 1976 was probably the least discussed phase in the literature. This phase is not normally considered as part of software methodology since it is performed most often by the user instead of

*It is worth remembering that the state of the art is defined as majority practice.

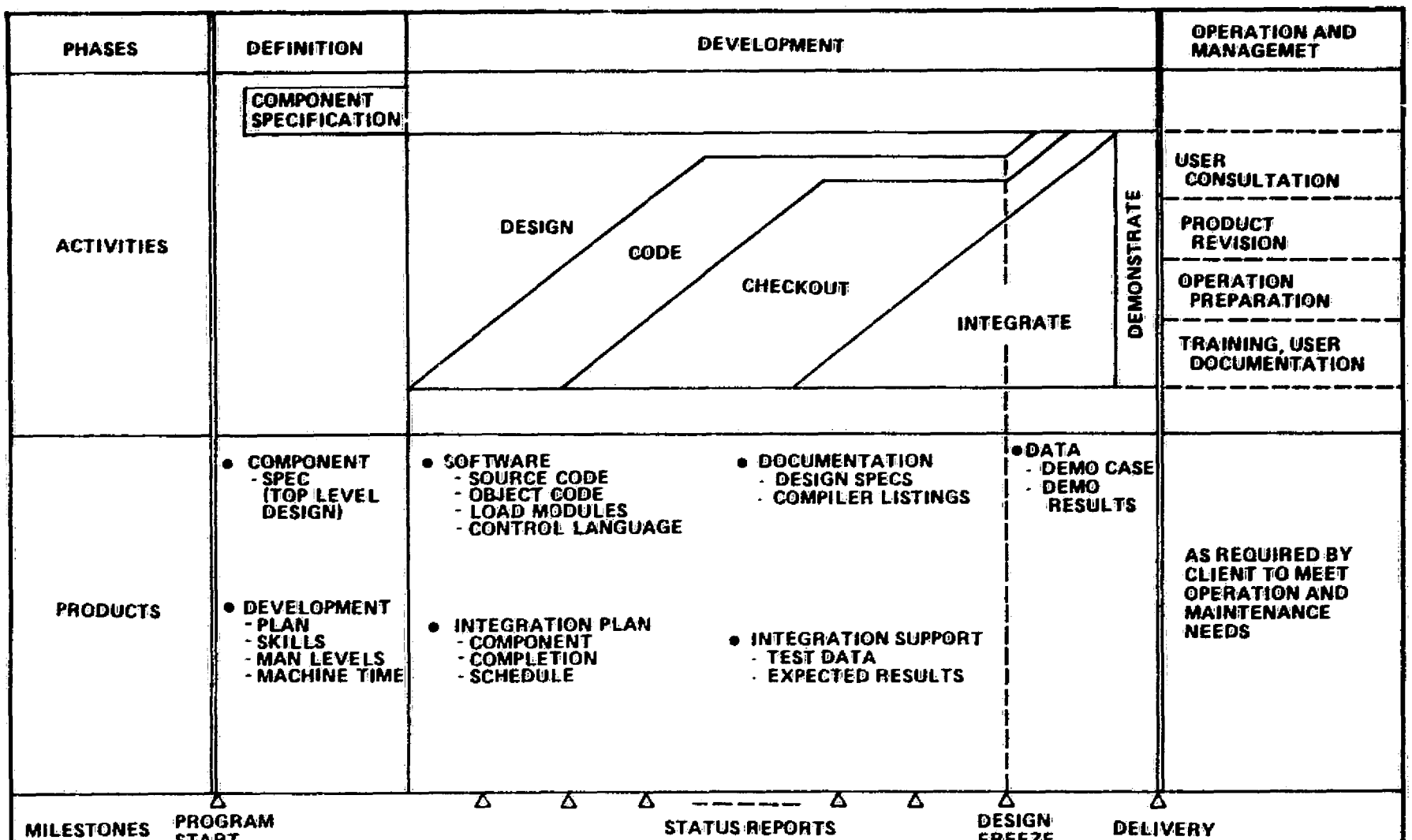


FIGURE 11.3.1.1.2-1. PROGRAMMING PRACTICES LIFE CYCLE - TRADITIONAL

the programmer. In fact, in combination with systems specification during the development phase, it is the root cause for virtually every problem that occurs in the Development Phase. Figure 11.3.1.2-1 (Ref. 11-53) graphically demonstrates this fact and shows how deficiencies that occur before the design activity of the Development Phase (all activities before DSARC II) lead to deficiencies through the rest of the software life cycle. This chart references military command and control systems and uses DOD milestones but is applicable to all types of applications programs.

Ideally, after initial concept formation by the user, both user and programmer should proceed through the Initation Phase and at least partly into systems specifications together.

Schwartz (Ref. 11-52) states which questions should be answered during this phase:

"1. What should be produced?
2. Should it be produced?
3. Can it be produced?"

He then states that the results of the phase (including systems specification) should include:

"1. A clear, detailed functional description of the total system, including the user interface with the system.

2. A reasonably accurate statement of budgets and schedules.

3. The resources available, required, and a plan for contingencies."

The state of the art falls considerably short of obtaining these results. Generally, instead of working with the development programmers, the user prefers to be disassociated, resulting in a virtual separation of Initiation Phase and the systems specification stage of development. The requirements are output in English (an ambiguous language) and generally fall short of systems considerations. There are no standards; user interfaces are frequently implied; requirements are in conflict or unnecessarily restrictive; and timing requirements are missing. Although the major method of acceptance will be testing, testing requirements are often omitted or are so vague as to have no specific meaning.

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

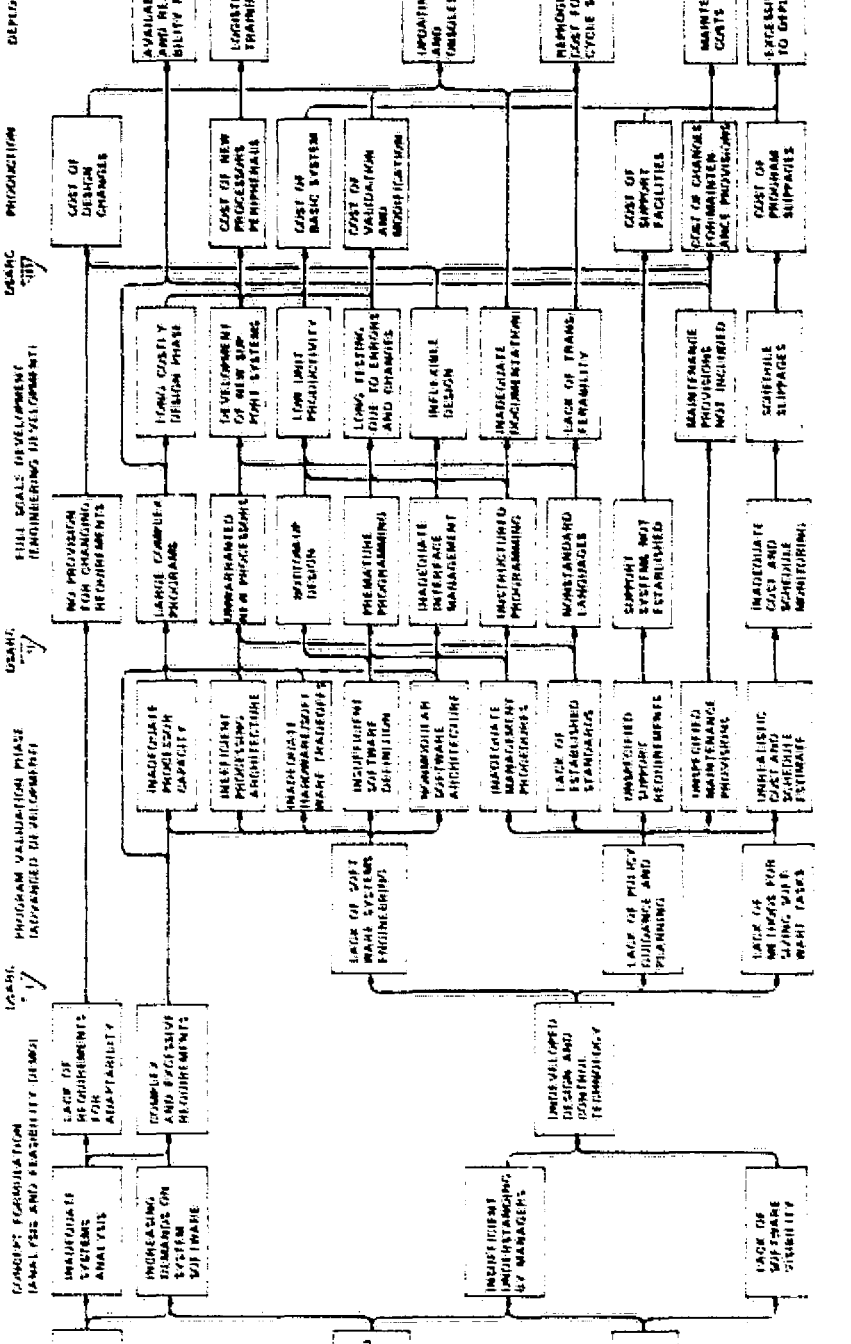


FIGURE 11.3.1.2-1. INTERRELATION OF SOFTWARE STUDY FINDINGS

It is with this start that most Development Phases begin. Yet it is still the state of the art that the Initiation Phase, which is the foundation for software development, is the least-understood phase in the software life cycle.

11.3.1.3 State of the Art as It Relates to the Development Phase - It should come as no surprise that software methodology is generally associated with the Development Phase of software. The idea that software methodology is associated with the entire life cycle is only beginning to find acceptance in the industry.

The state of the art of the Development Phase basically is the general state of the art already described for 1968. This section only summarizes the state of the art by subphases, addressing differences, and then addresses other aspects of the Development Phase as they relate to the state of the art.

The system specification phase as it should be performed was described in the previous subsection. In practice, it seldom receives the attention or time required to properly complete the specifications and more often is performed in near concurrence with the design phase. In fact, it is still common in the industry to reduce the Development Phase to just the remaining subphases by including system specifications as part of design.

The remaining phases are more distinguishable in today's Development Phase than in 1968 and more time is generally spent on design efforts. The changing emphasis on design can be seen by comparing the relative efforts on these phases (Table 11.3.1.3-1). The time spent on design is increasing, while the time spent on testing is generally decreasing. Table 11.3.1.3-1 is deceptive, however, in that it must not be assumed that all design is complete before coding starts. In practice, there is considerable overlap between phases and large projects still end up with all phases simultaneously active as in 1968. However, the process is better organized and designs are more complete.

TABLE 11.3.1.3-1. RELATIVE LEVELS OF EFFORT COMBINED FROM A NUMBER OF SOURCES

	ANALYSIS AND DESIGN (%)	CODING (%)	TESTING (%)
General Programs Before 1960	20	40	40
Spaceborne (Gemini, Saturn) (1960's)	34	20	46
OS/360 (1960's)	33	17	50
Modern Command and Control (1970's)	46	20	34

Design philosophy has been influenced by the technique of top-down design. Of all the methodologies proposed or tried in the last decade, this may well be the most accepted one. The technique revolves around starting with the specification and functionally describing the system on a general level, then increasingly defining more and more levels of sub-functions until the system is defined in detail. There is no widely accepted method of doing this, however, with both languages (process design languages) and graphics (functional flow diagrams) in use. There is no agreement as to when the functional breakdown should stop, hence software is frequently underdesigned and many critical decisions are still left for the coding phase. Too often the design reviews are scheduled too early in the project by both managers and users who want to see code output.

Design phases almost always end with a set of documents (not necessarily complete) describing the design, but there are still no standards and surprisingly no general agreement as to the contents of the documents or the level of detail. When the initial design stage is complete and coding starts (if it has not already), the design documents are quickly set aside, not maintained, and soon are unusable.

The coding phase proceeds much as it did in 1968. It is an interesting observation that the industry now believes in using some form of structured code, although in practice it seldom does. It is true, however, that the trend toward using higher-level languages is continuing

in the applications area and that the industry believes in "modularity" even though designs are seldom complete enough to ensure that modularity will be maintained through the testing stage. Coding styles still vary from programmer to programmer due to lack of standardization, and if standards are used in the beginning, they are likely to be relaxed in the rush to complete the program. Program commentary follows along the same lines as coding styles in general. One of the most significant defects in the current methodology is the lack of code traceability. Programmers are seldom required to acknowledge their code, yet it is a fundamental control principle that every line of code in every program be traceable to the author and that he be held responsible for it. This would help standards to be maintained and would prevent the partial incomplete "fixes" that so frequently occur during the coding and testing stages.

The coding stage has received the most attention of any stage in the Development Phase of software; however, its success is still dependent on individual programmers.

Testing is almost as neglected as the specifications stage. In fact, the scenario of testing presented for 1968 could easily be applied today. Test plans are seldom developed until late in the project and individual programmers are still responsible for developing test cases for their modules or routines before integrating them into the system. Initial test procedures still rely heavily on traces and dumps. Test records are not generally maintained. Integration testing still proceeds as in the 1968 scenario. In summary, with the exception of a few large programs where an outside contractor is responsible for software validation and verification, testing procedures have changed little in the last decade.

For the most part, software organizational structures have changed little since 1968. Programmers are nominally grouped around a leader, or on large programs, one of a number of leaders who is (are) nominally responsible for their area through all stages of the Development Phase. Some organizations have added a librarian to the project organization.

The librarian's function is to protect the baseline version of the program and perform program updates. Other organizational structures include the use of chief programmer teams and/or quality control groups. Chief programmer teams in practice operate much like the group leader concept above except that the chief programmer is held directly responsible for the design and all code produced by his group.

The increase of formal software quality control groups is a phenomenon that is discussed as a trend (Subsection 11.3.2.3) since quality control is the responsibility of the lead programmer (or chief programmer) for the majority of software organizations today and is practiced haphazardly. However, the increasing use of quality control groups indicates that, in 1978, the software industry is becoming concerned over the quality of the software it produces.

11.3.1.4 State of the Art as It Relates to the Operation and Maintenance Phase - The Maintenance Phase is the third and final phase in the software life cycle. It is also the phase where programming is most expensive. For example, it has been estimated that avionics maintenance costs can run 50 times (Ref. 11-54) per line of code more than the associated development costs. It was also pointed out, in Section 11, that 32 to 60% of all programming costs in the software life cycle occur during this phase. Some of the reasons for this activity are:

- The delivered software has "bugs" in it.
- The delivered software does not do exactly what it was supposed to do.
- Some of the requirements changed during the development phase.
- New uses were conceived, resulting in the need to enhance the software.
- Maintenance programmers discovered a "better" way to perform some function.
- The software must be transported to a new computer.

Maintenance programming methodologies are identical to those described for 1968. It is part of the state of the art that maintenance programmers feel that the program needs continual tuning and review. It is also part of the state of the art that the degree of configuration control exercised during this phase ranges from none at all to so much that even emergency error correction takes weeks to perform. Maintenance coding, however, is seldom performed with the thoroughness that was used during the Development Phase (with the possible exception of major modifications). The design stage and testing stage are generally informal and test only the correction. System testing is generally omitted. Thus the possibility of one "fix" causing more errors is high, ranging from estimates of one error per two fixes to two errors per fix — the latter suggesting why software seems to degenerate with age.

Although maintenance programming costs are high, there is little research or impetus to do anything to lower them. Since maintenance programming is performed after the development cycle, it must lag the state of the art and does by the 6 months to 4 years required to develop programs. All hopes for reducing costs in this phase rest on the hope that better quality, better structured, better documented, more reliable programs in the future will reduce the requirements for maintenance.

11.3.2 Trends in Software Methodology

11.3.2.1 General Trends in Software Methodology - Although the previous sections presented a gloomy picture of the state of the art, software methodology was not stagnant during the last decade, but instead industry took a wait-and-see attitude. In fact, starting in 1972 with the publishing of the paper about the New York Times Project and continuing today, methodologies are being subjected to experimentation and testing. However, many of the experimental projects fail due to the lack of understanding of the difficulties in monitoring software activities. A few successful projects are not generally accepted as proof of success due to the lack of meaningful software metrics.

For example, the New York Times Project used almost completely different methodology (this was described in Subsection 11.3.1.1.1) from the traditional methodology and claimed a phenomenal productivity of 35 source commands per day. It would have had far more impact if it had not been mentioned that the project had been carried out with a small team of very good programmers. The industry feels that any methodology would work with this type of group and therefore decides to wait for results from some large projects, a number of which are in progress. In summary, no individual techniques or methodologies have found general acceptance.

Research in methodology, particularly tools and testing techniques, has continued. Examples of the techniques and tools follow:

- Automatic Generation of Self-Metric Software (Ref. 11-55)* - The problem of software validation requires a deeper understanding of software behavior than merely "performance" analysis measured by conventional hardware or software monitors. Questions of performance and speed become secondary to the question of whether the software produces valid answers meeting problem requirements. A new measurement approach is self-metric instrumentation. It enables the user to study the internal structure and behavior of software systems. The self-metric instrumentation is an approach to the measurements of equipment utilization, instruction-type or resident-routine-type

*Although the research is attributed to the original author, this description is taken from Reference 11-43.

usage, program execution activity at the source-statement level, and program execution activity at the machine-language level. The measurement of program execution activity at the source-statement level is a technique whereby self-measurement capabilities are built into source-language programs. This can be accomplished by pre-processing the original program or augmenting the compiler to perform the necessary instrumentation. After the program is made self-metric, it is then executed and a subsequent post-processing step or a set of run-time routines then generates a report on actual program behavior. Two techniques have been experimentally used to implement sensors: 1) direct code insertion and 2) invocation of run-time routines. The direct code insertion appears to be faster in most cases, but the run-time routine is more flexible in that execution time measurements can be more easily altered.

The measurements performed allow examination of the internal structure of a software system rather than merely treating it as a "black box". While the testing process does not technically prove the correctness of the algorithms, it does allow observation of behavior with actual test data. From these observations, the effectiveness of the software verification process can be greatly improved.

An automated tool called Program Evaluation and Tester (PET) is developed according to this philosophy. PET consists of two major components: 1) a highly structured pre-processor that instruments the source program in a manner that makes it self-metric and 2) a post-processor to generate reports from the execution measurement data produced by the instrumented self-metric software. Current versions of PET are in ASA standard FORTRAN IV.

- FORTRAN Automatic Code Evaluation System (FACES) (Ref. 11-56)* - FACES is a tool designed to aid in analyzing and debugging FORTRAN programs. Its input is a FORTRAN program on which it performs analysis and reports correct parameter alignment between routines, COMMON block alignment, variable initialization, and the future and history of specified variables.

  The system is composed of two main parts: 1) the FORTRAN front-end, which gathers information about the input program, and 2) a set of routines organized as a diagnostic package that evaluate the information and print warning messages concerning actual or potential errors. The front-end scans and parses the input program, simultaneously

*Although the research is attributed to the original author, this description is taken from Reference 11-43.

gathering information about the source code. A graphical representation of each routine analyzed is formed and information concerning interfaces between routines is gathered and stored in tables. The second portion of FACES consists of diagnostic routines that analyze the information stored in the tables, looking for possible danger signs and specific problems. Each routine performed is chosen for execution by the user.

- Symbolic Execution (Refs. 11-57, 11-58, 11-59) - Symbolic execution is a method intended to generalize the program testing capabilities by replacing the specific test data values with symbolic names. Thus the output becomes a mathematical formula that is true for any values going through the path of a program. The symbolic execution actually decomposes an entire program into many different formulas, and each formula is the result of the symbolic input data traveling through one of the many paths of the program. Although one can verify the correctness of each path, no method has been proposed that can efficiently distinguish all the execution paths from nonexecution paths, since there is no way to eliminate nonexecution paths in an arbitrary program (Refs. 11-60, 11-61).

  The spectrum of software validation ranges from testing to program proving. Symbolic execution offers a happy compromise; it is somewhere in between the two extremes. Both program testing and program proving have severe problems in their validation process: program testing often draws the criticism that "testing can only prove the presence, never the absence, of bugs" (Ref. 11-62), while the techniques developed to prove program correctness are not easily implementable and the automatic theorem provers, which are essential in the program proving process, leave much to be desired (Ref. 11-63). This leaves symbolic execution as the brightest hope in software validation. Research in this area is predicted to achieve a certain degree of success during the 1980-1985 timeframe.

The industry has started to invest heavily in software tools, and this trend will continue as software tools continue to improve. Software tools may be defined as anything that automates some activity in the software life cycle. Most software tools are themselves software programs. Actually, the industry has almost always had some tools, such as assemblers and compilers, but generalized tool development and use grew slowly until the mid 1970's so that their use will not be significant until the early 1980's.

In the sixties, tools were created to support individual projects as part of the project and then were either discarded or delivered as part of the software (the latter case was rare). The realization that the same tools were being created over and over led to the realization that more generalized versions of the tools could be created and sold. By 1980, programmers will not be allowed to create their own tools unless they can prove to their managers that the required tool is not available on the market.

Today there is a specific tool in existence or under development for virtually every aspect of the software life cycle. There are tools that provide automatic documentation, flow charting, cost estimating, code validity checking, project reporting, language structuring, code standards checking, and timing and storage analysis. There are also tools* that are used to aid in design**, system specification, and requirements generation and analysis. With all the specific tools, it was inevitable that some group would try to package the tools together into one big tool. A number of such tools are under development (plus numerous systems oriented to special applications) under such names as Software Engineering System and The Software Factory. These facilities attempt to provide as much software project automation as is currently possible within the state of the art.

It should be noted that software libraries, though yet undeveloped, are potentially powerful tools, although tools suffer from many of the same problems as software libraries. The two major problems are lack of portability (although some tools provide limited portability) and language restrictions. Lack of portability is caused by the difference in computer architectures, instruction sets, and program compilers. Language restrictions occur as the tool is designed to operate on a specific language (i.e., a code syntax checking routine).

*A number of these tools were discussed from their language aspects in Section 11.2.

**One such tool is IORL, created by Teledyne Brown Engineering.

The strong interest in tools stems from two theoretical concepts: 1) tools automate, thus improving productivity, and 2) the enforced use of tools forces software discipline on undisciplined programmers. Management sees tools as a means to reduce both costs and risks and can readily justify their purchase. This provides the strongest evidence that the trend toward increased organization and discipline in software development will continue.

Starting in the late 1960's and early 1970's, DOD, alarmed by the increasing costs and unrealiability of its military software, undertook a series of studies that culminated in the 1976 issue of DOD Directive 5000.29: "Defense Systems Software Management Plan". Since DOD is the largest software consumer in the world, this document caused a number of trends. The directive, aimed at improving both DOD software acquisition management and industrial software management, covered the following policy areas:

- Requirements Validation and Risk Analysis (part of the Initiation Phase) - This policy generally expanded the areas to be included in the determination of requirements for new systems. Most significantly, it directed that software requirements be considered with hardware requirements (a systems approach) and that risk analysis, preliminary design hardware/software methodology, external interface control, security features, and life cycle planning be included.

- Configuration Management of Computer Resources - This policy directed that software acquisition be treated as a full configuration like the hardware. No longer would DOD procure software as a data item. The significance of this is that industry was told that it has to manage the software development in the same organized way that it managed hardware.

- Computer Resource Life Cycle Planning - This policy made it clear that planning was to be performed in terms of the life cycle of the system. Thus short cuts taken during the Development Phase were no longer acceptable if they would create problems during the Maintenance Phase.

- Support Software Deliverables - This policy made all supporting software, such as tools, deliverable so that it could be used during the Maintenance Phase.

- Milestone Definition and Attainment Criteria - This policy proclaimed that software projects would pass through stages in an orderly fashion and that its progress be monitored.

- Software Language Standardization and Control - This policy directed the services to agree on a limited number of higher-order languages and standardize them. A subsequent directive, DOD 5000.31, specified that the approved languages were CMS-2, SPL-1, TACPOL, J3 Jovial, J73 Jovial, ANSI COBOL, and ANSI FORTRAN. In addition, DOD launched an effort to create a single general-purpose language that could be used for all military systems. This lanugage (nicknamed DOD-1) is currently under development. It is based on PASCAL.

Most of the trends in software methodology today are based on engineering technology and represent a trend toward the overall development of a software engineering technology. The trend in industry is to move away from the poorly organized craft of the late sixties and toward a more organized methodology characterized by engineering principles. At the same time, the industry is giving hope that programming will become a science in the foreseeable future.

The movement toward a more organized methodology generally includes adoption of some or all of the top-down practices (design, code, test); however, the way they are employed varies from organization to organization and the rigor with which they are used from project to project within the organization

Figure 11.3.2.1-1 shows how one organization (Ref. 11-64) has organized their programming practices in the software life cycle. Observe that the design phase has no overlap with the coding phase (indicating the stress on design completeness) and the recognition of the need for complete planning early in the project. Like most other organizations, this one has adopted top-down techniques, but their implementation is far different from the structured programming implementation used (and subsequently adopted) by IBM on the New York Times Project. The mechanics of the methodology feature the following:

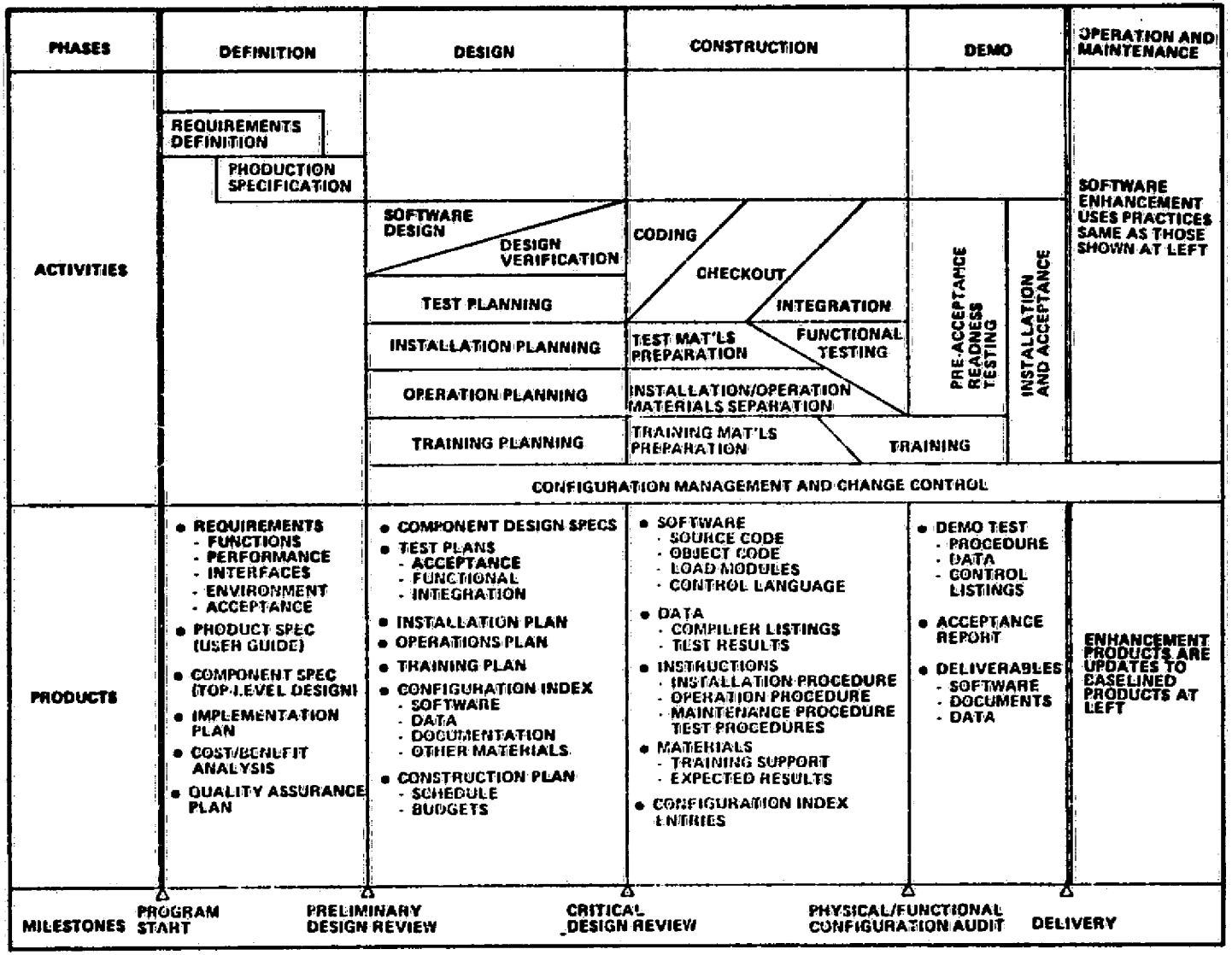


FIGURE 11.3.2.1-1. TRENDS IN PROGRAMMING PRACTICES LIFE CYCLE

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

- Program manager authority - The software project leader has management authority. He is given both technical and administrative responsibilities. He is responsible for meeting cost and schedule commitments, to deliver products and to satisfy quality criteria that he has responsibility to negotiate, and for discharging his obligations in a way that both satisfies the customer's needs and meets the company profit objectives. He has supervisory responsibility for making assignments, monitoring and directing task execution, and evaluating performance.

- Reviews - Use of formal reviews, keyed to phases of the development process.

- Unit development folder - A folder kept on each unit or module that serves to provide all records for that module. These folders contain the schedule, dates of task completion, requirements, design, test plans, test procedures, test results, interface descriptions, and code listings and verification.

- Design discipline and verification - Use of formal design and verification criteria. The design methodology is top-down to obtain the functional design. In addition, a construction plan is developed. The design is formally reviewed with the customer -- if possible before coding starts. Formal verification procedures are also used, such as structural walkthroughs, modularity analyses, data flow verifications, and compliance with requirements analysis.

- Support library and facilities - Each project establishes a controlled library as a location to record code, data, and other information related to the project. In addition, the library provides automated tools essential to the project, such as design aids, structured precompilers, and code scanners for adherence to standards.

- Phase testing - A formalized test procedure in which:

  - The project produces a test plan and test designs and secures customer concurrence on them.

  - Two types are developed: acceptance tests (formal demonstration) and functional tests (preferred if possible with customer's data).

  - The project produces test procedures and data cases and maintains them in a controlled library to facilitate test repeatability.

  - The procedures include pass/fail criteria.

  - The object of the tests is to ensure that the code correctly implemented the design.

- The tests are used as a basis to report test status and problems.

11.3.2.2 Specific Trends for the Initiation Phase - Some of the major trends for this phase were covered in the previous section. DOD Directive 5000.29 highlighted the problems of incomplete requirements and set in motion an attempt to solve them. All three services have completed drafts or their implementation documents. The trend toward better problem definition will be followed by the rest of the industry.

There is a strong trend in the industry to develop tools that make the processes of requirements specification during the Initiation Phase and system specification and design stages during the Development Phase all flow smoothly and unambiguously together. This is indicative of the general trend to reunite the user and the programmer in some fashion to develop a common understanding of the program under development. These tools all rely on some form of artificial language (see Subsection 11.2.1) as the basis for their input. It is typical of specification/design tools (see Table 11.3.2.2-1) that they utilize a data base; can (or will be able to) check for errors and inconsistencies; and produce graphic representations (except PDL), structure charts, data flows, or tables. Some of them can (or will be able to) run functional or analytical simulations and analyze the results.

These tools provide a means to produce meaningful specifications and designs, but most are only in some stage of development and only a few are designed for general applications. Yet they clearly point to the future.

11.3.2.3 Specific Trends for the Development Phase - Specific trends in software development are readily apparent in the literature and from the history of programming. One survey (Ref. 11-65) of 22 aerospace companies revealed that approximately 70% of the firms indicated that they had employed one or more of the precedures generally placed under the heading of Modern Programming Practices (MPP), although only 40% had appeared to have made a substantial commitment. There is a clear trend toward using at least some of the available software engineering techniques. This same survey listed the MPPs and asked which ones

TABLE 11.3.2.2-1. SOME EXAMPLES OF REQUIREMENTS/DESIGN SPECIFICATIONS TOOLS

NAME	DESCRIPTIVE NAME	TYPICAL APPLICATION AREAS	INPUT	DEVELOPER	GOVERNMENT SPONSOR
PSL/PSA	Problem Statement Language Problem Requirements Analyzer	Saturn IV	Text	Univ. Mich.	Air Force
SREP	System Requirements Engineering Project	Missile Systems	Diag/Text	TI, TRW	Army
HOS	Higher Order Software	Space Shuttle	Text/Code	Draper Lab	NASA, Navy, Army
PDL	Process Design Language	General	Text	C.S.C.	NASA
IORL	Input/Output Requirements Language	General	Computer Graphic Diag/Text	Teledyne Brown Engineering	I.R&D

were used*. The results, shown in Table 11.3.2.3-1, are enlightening. The five most used practices are relatively old practices (at least 10 years old) and are generally required on all Government contracts. It should also be noted that the diligence with which they were used is not available. On the other hand, a number of the practices associated with the New York Times Project are in moderate use. For projects employing MPPs, all of them except structured coding were employed more than 40%

TABLE 11.3.2.3-1. MODERN PROGRAMMING PRACTICE UTILIZATION (Ref. 11-63)

PRACTICE	PRECENTAGE REPORTED (%)
Program Manager Authority (Ref. 11-64)	78
Reviews (Ref. 11-64)	81
Unit Development Folder (Ref. 11-66)	37
Design Discipline and Verification (Ref. 11-64)	48
Program Modularity (Ref. 11-69)	70
Naming Conventions (Ref. 11-64)	56
Structured Form (Ref. 11-47)	15
Structured Walkthroughs (Ref. 11-45)	44
Structured Analysis (Ref. 11-67)	15
Structured Design (Ref. 11-67)	41
Chief Programmer Teams (Ref. 11-49)	48
HIPOs (Ref. 11-68)	11
Support Library and Facilities (Ref. 11-64)	44
Phase Testing (Ref. 11-64)	81
Configuration Management	78

Note: References listed provide the definitions used by the authors.

*The data are biased in that the authors asked for data from projects employing modern programming practices.

of the time. Remembering that the companies surveyed are part of the trend-setting industrial group and recalling the "7-year rule", there appears to be a general trend (although slow) to use the structured and associated software engineering technology. However, the transition to general usage will not occur until around 1982.

One structured technique, structured form (coding), does not appear to have wide acceptance. Its failure to gain acceptance can be explained: programmers do not like it and resist its use. In its pure form, structure coding limits the power of most higher-order languages to simple processing, one branch (IF...THEN...ELSE), and one loop (DO WHILE). Structured coding practices often utilize rigid formats relying on indentation to simplify the interpretation of the structure, require each unit to have one flow in and out, and impose standards on size of modules, commentary, etc. Today's programmers are too artistic to accept it. Today's managers, who are worried about loss of programmer morale and that users will become overly concerned with form, not content (hence increasing software costs), are reluctant to impose this discipline. Structured forms will not come into wide use in the immediate future. Instead, software engineering practice will concentrate on producing modular, well-organized code with simple interfaces between modules or routines. The organization and code in each subroutine will remain the prerogative of the programmer.

The trend toward tool use in the development phase is clear from the number of them marketed. In the survey, the most commonly used tools include automatic project monitors (46%), library monitors (30%), program structure analyzers (26%), and automatic test case generators (19%).

There is a trend toward on-line capabilities all through the Development Phase. In the sixties and through the middle seventies, all project using medium to large-scale computers were predicated on the restrictions imposed by limited access to the computer. This is because of the prevalence of the batch processing practice. One of the benefits of cheaper computer hardware and the introduction of the microprocessor

has been a decrease in cost of terminals. The terminal has the effect of returning the machine to the programmer. The programmer can enter his code, maintain working files (replacing card decks), and run his routines. The user can call up his programs and run them. It should be noted that most of the software tools are more effective if designed for on-line systems. This is particularly true of the requirements and design tools discussed in the previous section. The use of on-line tools is not without its problems, of which configuration control is the biggest, but the trend is clear: the use of on-line capabilities will grow until each person (or very small group) has his personal terminal.

The increased use of quality control groups is a recent event (occurring primarily during the last 4 years). However, software quality control is different from hardware quality control. Hardware quality control groups are generally independent groups who, through the use of statistical techniques, monitor the hardware to determine whether it meets its performance and quality specifications. During the engineering phase of a hardware development, the quality control group operates in the background to ensure that the hardware design and documents produced meet the required specifications. Quality control in hardware includes performance and reliability predictions.

Software quality control groups are almost primarily concerned with defect removal and to a lesser extent with enforcement of good programming practices. Thus software quality control is an extension of software management (the 1968 systems group renamed). The industry is becoming concerned with the quality of its software if for no other reason than the recognition that error-filled software is expensive to correct. It is 38 times (Ref. 11-69) more expensive to find and fix errors in the test stage than in the design stage, and 114 times more expensive after testing, yet almost half of all software errors are design errors. It is easy to believe the quality control thesis: "When control of the process improves, not only does the quality of the product increase but the cost of the product decreases". Quality control may

be performed informally or by a formal group, but the trend is toward specialists in a formal group.

11.3.2.4 Specific Trends for Operation and Maintenance - The trend in maintenance today is to wait for help in terms of clear requirements for better structured, better documented software from developers. However, in Subsection 11.3.1.4, six reasons were given for maintenance, and while it is true that better development of programs will greatly aid any maintenance programming effort, only two of the six reasons are directly related to the Development Phase. Although there should be a trend discouraging all but absolutely necessary changes, there is none.

One trend will begin as a result of the quality control movement. Maintenance programmers will be required to keep detailed error records (cause, type, etc.). Again, the trend that each line of code be traceable will start in this phase. This is important as all program errors are traditionally blamed on the developers.

11.3.3 Projected Developments in Software Methodology

11.3.3.1 General Forecasts - Forecasting software methodology is at best risky. The science of forecasting is supposed to be based on mathematical/statistical techniques using a data base accumulated in the past and tempered by outside events and their projected impacts on the statistically based trends.

Lacking the necessary metrics to perform a statistical trend analysis and a sufficient data base, it is necessary to choose a model to use as the basis for forecasting. Ideally, in the case of software methodology, the model should be in an area that has similar applications and yet be in a more advanced state of the art today. It has already been asserted that the trend in software methodology is toward employing software engineering technology. Thus the model chosen for forecasting the future of software engineering was digital engineering. Digital engineering is remarkable like software engineering in terms of the end product produced. Both deal literally in ones and zeros, AND's and OR's. Both produce complex end-products. Digital devices consisting of several printed wiring boards each with several LSI chips are not uncommon today. In fact, the computer is primarily a digital device and the software is but an extension that customizes the general circuitry.

New products are expensive in both disciplines. It is expensive to design a new chip or a new device just as it is expensive to produce software. However, the digital device can be mass-produced, so the engineering cost is spread over many copies. Unfortunately, software costs are always attributed to the first copy of the program.

Both digital engineering and programming (not software engineering) are about 25 to 30 years old. However, digital engineering is further advanced since it is a specialized form of electrical engineering, and unlike the programmer, the digital designer had the engineering discipline to guide him.

With the choice of a model, the next step is to determine where software engineering (S.E.) is relative to digital engineering (D.E.). Since digital engineering is about 25 years old and software engineering is about 10, it is reasonable to assume that D.E. is 15 years ahead of S.E. Actually, this is not a bad estimate from a historical view. Ten years (1961) after the beginning of D.E., the only chip available to the D.E. was the transistor. Even small-scale LSI was a dream of the future. Digital designs were cumbersome, and analog techniques were frequently chosen over digital. Digital designs were by necessity completed to the gate level. Libraries were crude, as was the engineering technology.

The forecasting model is good for roughly 15 years. The state of the art of digital engineering today could be used to predict the state of the art of S.E. in the early 1990's. Since this report is primarily concerned with forecasting to 1985, the "7-year rule" must be envoked to separate the state of the art and trends for that period.

Although most of the forecasts for software methodology are related to individual software life cycle phases and are given in the following sections, some general forecasts follow from the model and existing trends:

- The industry will move to an engineering and technician organization. The analyst/programmer of today's methodology who participates in all stages of the Development Phase will disappear. This change in organizational structure will be almost complete by 1985 for scientific applications and exist as a strong trend in other software areas. It will have no effect on programmer productivity.

- Use of tools will become commonplace by 1985.

- The industry will become data collection conscious and routinely use automatic data collection techniques to ensure accuracy. Universal software metrics will not yet exist, but some progress will be made.

- As a corollary to the third trend, code traceability will be a reality by 1985.

- The rate of increase in productivity will remain at 3% per annum as it has during the last decade.

The forecast that software engineering will move to the traditional engineer/technician organization is not in itself surprising, but the forecast that the organizational structure will change rapidly in the next 7 years is surprising in view of the fact that today analysts/programmers perform every phase. The prediction is based on economics. In view of rising salaries, managers will have to limit their senior people to the decision stages. The coding stage and a major portion of the testing stages could be performed by lower-paid (non-degreed) technicians under the direction of an engineer. By 1985, the use of programmer technicians will be common.

The prediction on tool use has the following rationale: The increasing on-line capabilities, strong current trends, and shortage of skilled personnel leave no alternative but to rely on any automated help that is available. Tool use is discussed again in the following subsections.

The forecasts on data collection are based on three current trends:

- Quality control groups want the data to refine cost-estimating procedures and to determine the quality of the software.
- DOD wants data to form a research base.
- Tools will be available to both provide project data to the manager while preserving it for other uses.

Virtually every manager will rely on automatic project monitoring tools.

It is unfortunate that the data collected will not be the same from installation to installation. However, experience will begin to point to useful metrics.

The final prediction is based on the absence of any trend that will result in a dramatic increase in productivity in the next 7 years. Areas could be developed now that would help. These areas are libraries and universal languages (discussed as a 1985 trend in the following sections). However, the industry is too immature to recognize these areas.

11.3.3.2 Initiation Phase Forecasts - This phase will be heavily influenced by the effects of DOD Directive 5000.29 over the next few years, and it is expected that complete analyses during this phase will become commonplace by 1985. Requirements analyses in the future will greatly benefit from the massive data collections that will be commonplace. By 1985, tools will permit some systems feasibility studies, and the need for these studies will be well-recognized.

The need for better communication between designers and developers will be well-recognized, and by 1985, attempts will be made to use existing requirements/design languages instead of the English specifications currently in use.

11.3.3.3 Development Phase Forecasts - By 1985, software engineering methodology will be well-defined for the total software life cycle. Even so, some of the methodology will only be used in research or experimental development. To forecast what this methodology will be like in 1985 and what trends will be prevalent, it is only necessary to examine the model's state of the art and then draw appropriate analogies. Digital engineering (see Table 11.3.3.3-1) utilizes the same phases and stages of development as software engineering except that the coding stage of software is called the model-building or fabrication stage by D.E.'s. Digital engineering relies heavily on the computer throughout its Initiation and Development Phases. In fact, without the automation and aid of the computer, digital engineering was sufficiently developed by 1970 to make automation feasible. Key factors leading to this state of development were:

- A common and well-defined specification and design language
- A common and well-defined development process
- Common cell, and later, chip libraries; a design philosophy to use them whenever possible.*

*They are nothing more than purchased parts.

TABLE 11.3.3.3-1. STATE OF THE ART IN DIGITAL ENGINEERING

PHASE/STAGE	ACTIVITIES	TOOLS	COMMENTS
Initiation (Requirements and System Specification)	1. Tradeoff Studies 2. Cost Studies 3. Feasibility Studies 4. Precise Specifications	1. Simulation to validate 2. Parts List catalogs 3. "Engineering" language 4. High Level Schematics	1. Specs - include input/output voltages, currents; logic levels; thresholds; tolerances on all components; physical properties, etc.
Design	1. Precise definition of the device using a hierarchal level of schematics down to the cell level 2. Printed wire boards (PWB) layout. 3. Photo plots to be used for semi-automatic manufacture	1. Simulation to validate 2. Computer aided design	1. Design emphasis is always on using as many "off the shelf" components as possible. 2. This is a highly automated process. 3. Cell libraries and sometime chip libraries are stored in the computer's files. 4. Layouts of chips and PWB's by draftsmen.
Model Building	1. Components are assembled on PWB by technician.		
Testing	1. Components tested 2. PWB's tested 3. Device tested	1. High speed digital test sets using test patterns derived during the design stage by simulation.	1. This chart stresses logic testing. There is also extensive environment testing and other types as specified by the requirements.

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

Unfortunately, none of these factors yet exist in software engineering methodology. But this field, which has automated every other field but its own, has the potential for even more efficient and automated procedures than digital engineering. Table 11.3.3.3-2 forecases the state of the art of software engineering by the year 2000. It is analogous to the state of the art of digital engineering today with one exception: the coding phase is itself largely automated (this is why 2000 was chosen instead of 1993)! The keys to this prediction are:

- The development of a universal specification/design language that can be understood by the computer
- A common component library.

These are the same factors that permitted the automation of digital engineering, since the common and well-defined development process is implied and enforced by the universal language.

Previously it was pointed out that there were only two known ways to increase programming productivity. Analysis of the techniques of digital engineering show there are two more, yet unexplored ways with unlimited potential: libraries and automatic code generation resulting from the universal language.

One of the biggest failures of current software methodology is in the use and maintenance of software libraries. After 25 years, only 10 to 20% of application program code (Ref. 11-1) is derived from existing code, and it is generally derived from the limited library supplied by the computer manufacturer. Industry libraries are generally outdated, poorly maintained, and poorly documented. In today's state of the art, no one seems to use or show concern for software libraries. Yet as has been shown, the slight increases in programmer productivity brought about by higher-level languages and increasingly better project controls have not come near meeting the demands of increasing program complexity and size. In fact, software threatens to become the limiting factor in the industry's growth. It is the state of the art that the industry is failing to recognize the potential increase in productivity

TABLE 11.3.3.3-2. SOFTWARE ENGINEERING - FORECAST FOR THE YEAR 2000

PHASE/STAGE	ACTIVITIES	TOOLS	COMMENTS
Initiation (Requirements and System Specification)	1. Tradeoff Studies 2. Cost studies 3. Feasibility studies 4. Precise specification in Universal Graphic Language	1. Simulation to validate 2. Accurate code estimates accounting only for unique new code only plus design complexity. 3. Computer aided data base	1. All tools are based on computer analysis of the universal language.
Design	1. Precise definition of program using Graphic Hierarchal representation which includes functional diagrams, data tables, and data flow diagrams	1. Simulation to validate each step of design. 2. Computer aided design	1. Design emphasis is always on using as many "off the shelf" components as possible. 2. Component libraries stored in computer files.
Code	1. Code generation to convert basic program and all library components to ANSI Higher level Language 2. Special coding in ANSI HOL by programmer technicians	1. Computer generated code (directly from the Universal Language)	1. Choice of HOL is to provide as much portability as possible.
Test	1. Testing has been continuous. 2. Hand coded components tested 3. Program tested	1. Automated testing	1. Tests used for this phase were automatically derived from Universal Design Language.

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

that is possible from greater use of libraries. If productivity were seen in terms of the ratio of total application program size (machine instructions) to total person-project days, then the industry would have a generalized productivity metric that more realistically reflects its capabilities to meet future requirements.*

It is not unrealistic to expect applications programs to approach 80% library code utilization if good libraries exist. If this goal were realized, programmer productivity would increase by a factor of four, more than over the first 25 years!

Before libraries can reach their potential, a number of attitudes must change and existing problems be solved:

- Management must recognize that software is capital property and must initiate inventory controls.
- Users must develop techniques for tailoring their requirements around maximum library use.
- Computer manufacturers must determine how to best standardize "kernels" of the hardware (instruction sets) (possibly with firmware) to make software totally portable.
- The software industry must stop creating new coding languages.
- Compilers and languages must be standardized.
- A national clearing house must be established as a medium for buying and selling library components.

The industry has the capabilities to develop the necessary universal language today (IORL has just this potential), and code generation and libraries are already used on a limited scale. If the industry knew where it was going, the forecast for the year 2000 could become a forecast for the late 1980's or even for 1985. It is unfortunate that software methodology will fall considerably short since the industry has neither the desire nor the impetus to accomplish such a radical change in such a short time. It is fundamental, however, that

*The definition of productivity omits the operating system directly but includes code generated by the compiler and would include any section of system software code envoked by the application program such as envoked by Input/Output commands and calls to the system library.

with or without knowing it, the software industry will move toward this forecast for the year 2000.

Based on current trends, the following forecasts can be made for the Development Phase for 1985:

- Structured techniques for specification, design, and test design will be commonplace. Computer tools in the forms of on-line capabilities will greatly aid these processes.

- The steps involved in producing software will be formally defined, leading to more disciplined programming practices.

- Modular, well-organized code will be the rule, but structured code as it is defined today will have limited use. Code will be written by programmer technicians, not by software engineers.

- Chief programmer teams will give way to the trends toward software engineers and programmer technicians. the reasons for this were discussed in Subsection 11.3.3.1.

- No new programming language will make any great inroads on the languages in use today. (Corollary: DOD-1 will be used only when required; it does not fit into the long-range model).

- Testing will be performed by both technicians and engineers. However, this area will still lack the sophistication of other stages in the Development Phase.

- Automatic project monitoring to facilitate project management will be commonplace.

- Formal quality control groups will become a standard part of all projects. They will be active throughout all stages of the Development Phase and will start to make attempts to predict software reliability. It is from these people and their motivations that the beginnings of automatic data gathering will start and future trends will be conceived. (Corollary: Configuration Control, partially automated, will become a reality.)

11.3.3.4 Operation and Maintenance Phase Forecasts - As in the past, methodology in the Operation and Maintenance Phase will lag the changes in the methodology of the Development Phase, although it generally will follow similar lines. By 1985, the high costs of software operation and maintenance should begin to show a downward trend due to a number of factors:

- Better organized and documented software will make changes less complex.
- Clearer requirements will reduce the need for changes.
- The increased use of packaged software will reduce the general tendency to modify it. Users will learn to accept software as delivered.
- Users will begin to expect software to be warranted (some already do!) and produced for a fixed cost. This will result in both increased reliability and decreased tendencies by users to modify the software.
- Line-by-line code responsibility will be common and result in programmers accepting stringent configuration management and test requirements.
- Major modifications will be subject to the same methodology as used in the Development Phase.

In summary, the payoff of software engineering methodology does not lie in lower-cost software development (although there is some evidence that software development will cost less by using modern methodology) but in lower costs for maintenance, hence lower costs for the complete life cycle of the software.

REFERENCES - SECTION 11

11-1 NASA Software Management Guidelines (Draft), July 1976

11-2 Robert C. Kendall, "Management Perspectives on Programs, Programming and Productivity", Proceedings of the AIIE "Software" Conference, December 12-14, 1977

11-3 Jacques S. Gunsler, "Software Management in the Department of Defense", Abridged Proceedings from The Software Management Conference, Los Angeles Section of the AIAA in cooperation with the ACM and IEEE Computer Society, First Series, 1976

11-4 B. W. Boehm, "Software Engineering", IEEE Transactions on Computers, December 1976

11-5 J. R. Bright, "The Process of Technological Innovation", A Guide to Practical Technological Forecasting, Edited by J. R. Bright and M. E. F. Schoeman, Prentice-Hall, Englewood Cliffs, New Jersey, pp. 3-12

11-6 D. Good, R. Landon, and W. Bledsoe, "An Interactive Program Verification System", Proceedings of 1975 International Conference on Reliable Software, April 1975

11-7 A. R. Soxena and T. H. Bredt, "Verification of a Monitor Specification", Proceedings Second International Conference of Software Engineering, October 1976

11-8 D. I. Good, "Provable Programming", Proceedings of 1975 International Conference on Reliable Software, April 1975

11-9 E. S. Cohen, "Protection in the HYDRA Operating System", Operating Systems Review, Special Issue, November 1975

11-10 T. A. Linden, "Operating Systems Structures to Support Security and Reliable Software", Computing Surveys, Special Issue: Reliable Software II - Fault-Tolerant Software, December 1976

11-11 S. T. Walker, "A Certified Multilevel Secure Minicomputer Operating System", Signal, November/December 1977

11-12 G. E. Brown, R. Eckhouse, Jr., and J. Eslabrook, "Operating System Enhancement through Firmware", Proceedings Tenth Annual Workshop on Microprogramming, 1977

11-13 "Status Report of the Graphic Standards Planning Committee of ACM/SIGGRAPH", Computer Graphics Quarterly Report, Vol. 11, No. 3, Fall 1977

11-14 William E. Shea, "DOD-1, A Common Language", ISRAD in Touch, Vol. 2, No. 1, February 1978

11-15 C. V. Ramamoorthy et al., "Software Requirements and Specifications: Status and Perspectives", Engineering Research Recommendations, Draft Appendix A, August 12, 1977

11-16 B. W. Boehm, "Software and Its Impact: A Quantitative Assessment", Datamation, pp. 48-59, May 1973

11-17 T. A. Dolotta et al., Data Processing in 1980-1985, John Wiley and Sons, 1976

11-18 T. A. Thayer et al., "Software Reliability Study: Final Technical Report", TRW Report 75-2266-1.9-5, March 1976

11-19 B. W. Boehm, "Some Steps Toward Formal and Automated Aids to Software Requirements Analysis and Design", IFIPS 74, pp. 192-197, 1974

11-20 F. C. Ihrer, "The Projection of Computer Performance Thru Simulation", Comress, Third Addition, pp. 23-28, August 1966

11-21 D. Vedder, "GRAFPAC User's Manual", BTL Memo, July 16, 1970

11-22 F. M. Trapnell, "A Systematic Approach to the Development of System Programs", Spring Joint Computer Conference, pp. 411-418, 1969

11-23 Igor F. Stravinsky, Poetics of Music in the Form of Six Lessons, Cambridge, Harvard University Press, 1947

11-24 "HIPO: Design Aid and Documentation Tool", First Edition, IBM, April 1973

11-25 M. Hamilton and S. Zeldin, "Higher Order Software - A Methodology for Defining Software", IEEE Transactions on Software Engineering, Vol. SE-2, No. 1, pp. 9-32, March 1976

11-26 "GSFC NAVPAK Design Specification Languages Study", Final Report, Contract No. NAS5-23428, Martin Marietta Corp., October 1977

11-27 C. R. Everhart, "IORL Analysts Users' Manual", Section 1 - Syntax, Teledyne Brown Engineering, May 1977

11-28 C. R. Everhart, "IORL Analysts Users' Manual", Section 2 - Semantics, Teledyne Brown Engineering, May 1977

11-29 C. R. Everhart, "Introduction to IORL", Teledyne Brown Engineering, May 1977

11-30 S. C. Wheeler and C. R. Everhart, "IORL Operator Users' Manual" Teledyne Brown Engineering, May 1977

11-31 R. Bieri, "Telemetry Computation Branch, Quality Assurance Procedures, Program Design Language", Contract NAS 5-11999, Computer Sciences Corporation, October 27, 1976

11-32 PDL a Tool for Software Design, Caine, Farber & Gordon, Inc.

11-33 D. Teichroew et al., "Problem Statement Language", PSL Language Reference Manual, University of Michigan, January 1974

11-34 D. Teichroew et al., "An Introduction to PSL/PSA", ISDOS Working Paper No. 86, University of Michigan, March 1974

11-35 M. E. Dyer et al., "REVS User Manual", SREP Final Report, Volume II, TRW, August 1, 1977

11-36 D. T. Ross and J. W. Brackett, "An Approach to Structured Analysis", Computer Decisions, pp. 40-44, September 1976

11-37 Henry Kleine, "Software Design and Documentation Language", JPL Publication 77-24, National Aeronautics and Space Administration, July 1, 1977

11-38 "NASA Software Specification and Evaluation System Design", Final Report, Part I - Software Verification/Validation Techniques, Contract NAS8-31379, Science Applications, Inc., March 19, 1976

11-39 C. R. Everhart, "SAFEGUARD System - Requirement Specification Language", BTL Memo, September 3, 1969

11-40 J. W. Hurst, "The Problem Oriented System"

11-41 T. A. Dolotta et al., Data Processing in 1980-1985, John Wiley and Sons, 1976

11-42 D. L. Parnas, "A Course on Software Engineering", Proceedings of ACM SIGCSE Second Technical Symposium, March 1972

11-43 Pei Hsia, Survey and Recommendations on Verification Methodologies to Improve Software Quality, Sponsored by Computer Sciences Division of Computer Sciences Corporation

11-44 H. Wirth, "Program Development by Stepwise Refinement", CACM, pp. 221-222, April 1971

11-45 G. M. Weinberg, The Psychology of Computer Programming, Van Hostrand Reinhold, 1971

11-46 Corrado Böhm and Giuseppe Socopini, "Flow Diagrams, Turing Machines, and Languages with only Two Formation Rules", Communications of Association for Computing Machinery, Vol. 9, pp. 366-371, 1966

11-47 E. W. Dijkstra, Notes on Structured Programming, August 1969

11-48 Harlan D. Mills, "Mathematical Foundations for Structured Programming", IBM, Gaithersburg, Maryland, Rep. FSC 71-6012, February 1972

11-49 F. T. Baker, "Chief Programmer Team Management of Production Programming", IBM Systems Journal, Vol. 11, No. 1, pp. 56-73, 1972

11-50 Barry W. Boehm, "The High Cost of Software", Practical Strategies for Developing Large Software Systems, Edited by Ellis Horwitz, Addison-Wesley Publishing Company, Reading, Massachusetts, 1975

11-51 Wernier L. Frank, "The Ten Great Myths of Software", a speech presented at the AIIE "Software" Conference, December 12-14, 1977

11-52 Jules I. Schwartz, "Construction of Software: Problems and Practicalities", Practical Strategies for Developing Large Software Systems, Edited by Ellis Horwitz, Addison-Wesley Publishing Company, Reading, Massachusetts, 1975

11-53 "DOD Weapon Systems Software Management Study", Johns Hopkins University Applied Physics Laboratory Technical Report, May 1975

11-54 W. L. Trainor, "Software - From Satan to Savior", United States Air Force Avionics Laboratory, Wright-Patterson AFB, Ohio, presented at National Aerospace Electronics Conference, May 1973

11-55 L. G. Stuki, "Automatic Generation of Self-Metric Software", 1973 IEEE Symposium on Computer Software Reliability, New York City, 30 April - 2 May 1973

11-56 J. C. Browne, R. Kleir, T. Davis, M. Henneman, A. Haller, and G. L. Lasseter, "Specifications and Programs for Computer Software Validation Information Research Associates", Contract NAS8-28084, November 1973

11-57 W. E. Howden, "Reliability of the Path Analysis Testing Strategy", IEEE Transactions on Software Engineering, SE-2, No. 3, September 1976

11-58 J. L. King, "A New Approach to Program Testing", Proceedings of 1975 International Conference on Reliable Software, April 1975

11-59 R. S. Boyer, B. Elspas, and K. N. Levitt, "SELECT, A Formal System for Testing and Debugging Programs", Proceedings of 1975 International Conference on Reliable Software, April 1975

11-60 L. J. Osterweil, "A Proposal for an Integrated Testing System for Computer Programs", University of Colorado, Computer Science Department, CU-CS-093-76

11-61 C. V. Ramamoorthy, S. F. Ho, and W. T. Chen, "On the Automatic Generation of Program Test Data", IEEE Transactions on Software Engineering, SE-L, No. 4, December 1976

11-62 E. W. Dijkstra, "Structured Programming", Software Engineering Techniques, NATO Science Conference Record, Edited by J. N. Buxton and B. Randell, 1970

11-63 B. Elspas, K. Lebitt, R. Waldinger, and A. Waksman, "An Assessment of Techniques for Program Correctness", Computing Surveys, 4(2), June 1972

11-64 Rachel K. E. Black, "BCS Software Production Data", BCS Report F30602-76-C-0174, Prepared for Air Force Rome Air Development Center, Boeing Computer Services, Inc., Seatlle, 1977

11-65 Richard H. Thayer and John H. Lehman, "Software Engineering Project Management: A State of the Art Report", A Collection of Papers, AIAA/NASA/IEEE/ACM Computers in Aerospace Conference, Los Angeles, California, October 31 - November 2, 1977

11-66 F. S. Ingrassia, "The Unit Development Folder (UDF): An Effective Management Tool for Software Development", TRW-SS-76-11, TRW, Inc., October 1976

11-67 Edward Yourdon, How to Manage Structured Programming, Yourdon, Inc., New York, 1976

11-68 "HIPO - A design Aid and Document Technique", IBM Installation Manual, GC20-1851-1, IBM, May 1975

11-69 James Frame, Keynote Address of the AIIE "Software" Conference, December, 1977

APPENDIX A. SPACE-TO-GROUND COMMUNICATIONS ELEMENTS

The researching of state-of-the-art characteristics for active devices in the microwave region produced a number of lists for the different types of devices. These lists contained useful information about the characteristics of currently available solid-state devices, including frequency, power, gain, efficiency, noise figure, and the device manufacturer. For reasons of completeness, these listings are presented in Tables A-1 through A-6. Table A-1 presents device information for a number of FET devices. Table A-2 covers the same type of information for bipolar transistors. Table A-3 lists information on TRAPATTS and IMPATTS in the 3-to-4-GHz range, and Table A-4 covers these devices for the 8-to-10-GHz range. Table A-5 presents information on IMPATT diodes for frequencies above 12 GHz. The characteristics for a number of small-signal FET devices suitable for microwave receivers are listed in Table A-6.

TABLE A-1. STATE OF THE ART IN FET DEVICES

FREQUENCY (GHz)	POWER (W)	POWER GAIN (dB)	EFFICIENCY (%)	MATERIAL	SOURCE
4	0.26	9.6	68	GaAs	RCA
4	3	6	47	GaAs	Fujitsu
5	0.25	12	22	GaAs	Plessey
6	2.4	5	34	GaAs	Fujitsu
8	1	7.0	-	GaAs	MSC
8	-	20	-	GaAs	Plessey
8	4.1	4	31	GaAs	Texas Instruments
8 to 10	0.7	6	46	GaAs	Texas Instruments
10	-	10	-	GaAs	Plessey
12	0.5	6	15	GaAs	RCA
15	0.45	5	12.5	GaAs	RCA
22	0.14	4.5	9	GaAs	RCA
47	0.1	-	2.2	InP	Plessey

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR

TABLE A-2. STATE OF THE ART IN BIPOLAR DEVICES

FREQUENCY (GHz)	POWER (W)	POWER GAIN (dB)	EFFICIENCY (%)	SOURCE
1	40	9	65	Typical
2	24	7	45	Typical
2	40	8	50	TRW
3	8	7	45	Typical
4	7	-	-	Texas Instruments
6	1.5	4	36	Hewlett-Packard
8	1.5	-	-	Texas Instruments
10	1	4.5	20	Texas Instruments

TABLE A-3. STATE OF THE ART IN 3-TO-4-GHz TRAPATT AND IMPATT DEVICES

DEVICE	PEAK P_o (W)	PW (μsec)	DUTY CYCLE (%)	1 dB BW (%)	POWER GAIN (dB)	EFFICIENCY (%)	SOURCE
TRAPATTS	240*	0.2	0.1	8.3	4.8	29	Hughes
	195	Narrow	-	Narrow	6.3	20	RCA
	98	0.2	0.1	8.3	6	30	Hughes
	82	10	?	8	4.9	15	RCA
	55	50	1	5	5	24	Hughes/NRL
	45	50	0.001	10	7	21	Sperry
	3	CW	100	Oscillator	-	20	BTL
IMPATTS	21 (3 diodes)	CW	100	Narrow	Locked Oscillator Locking Power = 3.5 W	12 to 13	BTL
	12.1	CW	100	Oscillator	-	21	Lincoln Laboratories
	3.4	CW	100	Oscillator	-	37	Lincoln Laboratories

*Two devices hybrid combined.

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

TABLE A-4. STATE OF THE ART IN 8-TO-10-GHz TRAPATT AND IMPATT DEVICES

DEVICE	MATERIAL AND TUBE DEVICE	PEAK P_o (W)	PW (μsec)	DUTY CYCLE (%)	BW (%)	POWER GAIN (dB)	EFFICIENCY (%)	SOURCE
TRAPATTS		27	1	0.1	Oscillator	-	42.5	Hughes
		10	0.25	0.1	10	6	12	Sperry
		6	CW	100	Oscillator	-	24	Hughes
		5	CW	100	5.5	5	16.4	Hughes
IMPATTS	Si DD	16	0.8	25	Oscillator	-	12.3	Hewlett-Packard
	GaAs HI-LO	12.8	0.8	25	Oscillator	-	25	Varian
	GaAs SB Read	8	CW	100	Oscillator	-	35	Raytheon
	GaAs Flat Profile	5	CW	100	7	25 (4 Stages)	10	Typical
	GaAs SB Read	4.5	CW	100	6.5	4.5	22	Raytheon
	GaAs HI-LO	2.9	CW	100	Oscillator	-	18.8	Varian

TABLE A-5. STATE OF THE ART IN IMPATT DEVICES ABOVE 12 GHz

MATERIAL	TYPE*	FREQUENCY (GHz)	PEAK P_o (W)	PW (μsec)	BW (%)	POWER GAIN (dB)	EFFICIENCY (%)	SOURCE
Silicon	DD	14.6	0.81	CW	Oscillator	-	15.1	Microwave Institute
	DD	16.5	11.0	0.8	Oscillator	-	14	Hewlett-Packard
	DD	17	10.0	10	600 MHz	20	-	Martin Marietta
	DD	29.39	1.2	CW	Oscillator	-	10	Hughes
	DD	39	11.0	0.35	Oscillator	-	10	Hughes
	DD	55	1.6	CW	Oscillator	-	11.5	Fujitsu
	DD	60	1.0	CW	Up to 6 GHz	9 (2 Stages)	2	Hughes
	DD	66	0.7	CW	Oscillator	-	6	Hughes
	DD	92	0.38	CW	Oscillator	-	12.5	GTI
	SD	94	0.1	CW	2 GHz	4	2	Hughes
	DD	140	0.72	0.3	Oscillator	-	4	Hughes
	DD	170	0.03	CW	Oscillator	-	2	Hughes
	SD	185	0.08	CW	Oscillator	-	2.3	Nippon T&T
	SD	202	0.05	CW	Oscillator	-	1.3	Nippon T&T
	DD	212	0.209	0.05	Oscillator	-	1.5	Hughes
	DD	215	0.52	-	Oscillator	-	2.6	Hughes
	SD	285	0.008	CW	Oscillator	-	0.35	Nippon T&T
	SD	301	0.0012	CW	Oscillator	-	-	Nippon T&T
GaAs	SB Read	13.7	3.2	CW	Oscillator	-	24	Raytheon
	DD	21	1.2	CW	Oscillator	-	15.6	Hitachi
	SD	34.8	0.7	CW	Oscillator	-	12.4	RCA
	SD**	37.5	0.5	CW	Oscillator	-	9.6	Lincoln Laboratories
	SB	51	0.2	CW	Oscillator	-	11.0	Hitachi
	SB	53	0.4	CW	Oscillator	-	5.2	Hitachi

* DD - Double Drift; SD - Single Drift; SB - Schottky Barrier

**SD Ion Imp.

TABLE A-6. SMALL-SIGNAL FET CHARACTERISTICS AND COST
(EXTRACTED WITH MODIFICATIONS FROM REF. 5-4)

MANUFACTURER OR U.S. REP.	MODEL	GATE LENGTH (μm)	NOISE FIGURE (dB min.)	TEST FREQUENCY (GHz)	GAIN (dB)		RF POWER OUTPUT (dBm)	SMALL QUANTITY PRICE ($)
					ASSOC.	MAG		
Avtech Industries 825 Stewart Dr. Sunnyvale, CA 94086 (408) 732-0880	AFT2000	0.75	2.2	8	9.0	12.0	10.0	$ 60 (10-49)
Dexcel, Inc. 2285-C Martin Ave. Santa Clara, CA 95050 (408) 244-9833	DXL-502	0.50	2.5	7	7.0	9.0	10.0	$ 95 (1-24)
Fujitsu America, Inc. 1135 E. Janis St. Carson, CA 90746 (213) 636-0858	FSC02	1.00	3.5	8	6.5	12.0	-	$250
Hewlett-Packard Company 1501 Page Mill Rd. Palo Alto, CA 94304 (415) 493-1501	HFET-1000	1.00	3.6	10	6.9	11.0	14.5	$135
Hitachi-Denshi 58-25 Brooklyn/ Queens Expressway Woodside, NY 11377 (212) 898-1261	HCRL-85	1.00	3.0	8	-	11.0	10.0	$ 60
	HCRL-87	0.50	2.5	8	-	13.0	10.0	$160
NEC (Nippon Electric) Calif. Eastern Labs One Edwards Ct. Burlingame, CA 94010 (415) 342-7744	NE244	1.00	2.3	8	10.5	12.0	8.5	$ 75 (10-24)
	NE388	0.50	2.7	12	8.0	12.0	8.5	$125
	NE463	1.0/1.0 Dual	3.2	8	14.0	15.0	8.5	$ 90
	-	-	1.55	4	-	30	-	-
Plessey Optoelectronics and Microwave 1674 McGraw Ave. Irvine, CA 92705 (714) 540-9945	GAT-5	0.80	2.5	8	10.0	12.0	13.0	$105
			3.0	10	8.0	10.0	10.0	
	GAT-3	-	3.5	12	-	5.8	-	-
	GAT-3	-	1.2	2	-	12	-	-

REPRODUCIBILITY OF THE ORIGINAL PAGE IS POOR